

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

Volume VIII

1948

917 FIFTEENTH ST., N.W., • WASHINGTON 5, D. C.

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

EDITOR

G. FREDERIC KUDER Duke University

ASSOCIATE EDITORS

DOROTHY C. ADKINS University of North Carolina
FORREST A. KINGSBURY University of Chicago
FRED MCKINNEY, Editorial Representative of the American College
Personnel Association University of Missouri
M. W. RICHARDSON Richardson, Bellows, Henry and Co.

BOARD OF COOPERATING EDITORS

JOHN G. DARLEY
University of Minnesota
HAROLD A. EDGERTON
Richardson, Bellows, Henry and Co.
MAX D. ENGELHART
Chicago City Junior Colleges
E. B. GREENE
United States Employment Service
J. P. GUILFORD
University of Southern California
E. F. LINDQUIST
State University of Iowa
CHARLES I. MOSIER
Personnel Research Section, A.G.O.
P. J. RULON
Harvard University

DAVID SEGEL
U. S. Office of Education
C. L. SHARTLE
Ohio State University
H. C. TAYLOR
*The W. E. Upjohn Institute for
Community Research*
THELMA G. THURSTONE
University of Chicago
HERBERT A. TOOPS
Ohio State University
E. G. WILLIAMSON
University of Minnesota
BEN D. WOOD
Columbia University
JOHN R. YALE
Science Research Associates

This journal is open to: (1) discussions of problems in the field of the measurement of individual differences, (2) reports of research on the development and use of tests and measurements in education, industry, and government, (3) descriptions of testing programs being used for various purposes, and (4) miscellaneous notes pertinent to the measurement field, such as suggestions of new types of items or improved methods of treating test data. Contributors receive one hundred reprints of their articles without charge. Manuscripts should be sent to G. Frederic Kuder, Box 6907, College Station, Durham, North Carolina. Writers are requested to include a biographical sketch with each manuscript, following the style of the section on contributors published in each issue.

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT is published quarterly, one volume per calendar year, at Mount Royal and Guilford Avenues, Baltimore 2, Maryland and 917 Fifteenth St., N. W., Washington 5, D. C. Entered as second class matter August 16, 1948, at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Subscription rate, \$5.00 a year, domestic and foreign. Single copies, \$1.50, with the exception of Volume VII, No. 3 (in two parts) and Volume VIII, No. 3 (in two parts), for which the price is \$2.25. Back volumes. Volumes V (1945), VI (1946), VII (1947) and VIII (1948), \$6.00 each. Volumes I through IV are available in a small-print edition at \$3.00 per volume (paper bound).

Orders should be sent to EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 917 Fifteenth St., N. W., Washington 5, D. C.

Copyright, 1948 by G. Frederic Kuder

INDEX FOR VOLUME VIII

<i>Adams, Sidney (with Milton M. Mandell)</i>	
SELECTION OF PHYSICAL SCIENTISTS	575
<i>Anderson, Gordon V.</i>	
PROFESSIONAL STANDARDS AND TRAINING FOR COLLEGE	
PERSONNEL WORKERS	451
<i>Baier, Donald E.</i>	
SELECTION AND EVALUATION OF WEST POINT CADETS	193
<i>Bennett, George K. (with Jerome E. Doppelt)</i>	
THE EVALUATION OF PAIRS OF TESTS FOR GUIDANCE USE	319
<i>Bixler, Ray H.</i>	
COUNSELING. ECLECTIC OR SYSTEMATIC?	211
<i>Brayfield, Arthur H.</i>	
PUTTING OCCUPATIONAL INFORMATION ACROSS.	485
<i>Butler, John M.</i>	
ON THE ROLE OF DIRECTIVE AND NON-DIRECTIVE TECH-	
NIQUES IN THE COUNSELING PROCESS	201
<i>Canfield, A. A. (with Neil D. Warren)</i>	
AN OPTOMETRIC APTITUDE TEST.	183
<i>Cannes, Earl F. (with Francis P. Robinson)</i>	
THE ROLE OF CLIENT TALK IN THE COUNSELING INTERVIEW.	635
<i>Chalfen, Leo</i>	
PRACTICAL SUGGESTIONS FOR IMPROVING TESTING PROCE-	
DURES	49
<i>Clark, Jerry H.</i>	
INTELLIGENCE TEST RESULTS OBTAINED FROM A SPECIFIC	
TYPE OF ARMY A.W.O.L.	677
<i>Clark, Kenneth E. (with Philip H. Kriedt)</i>	
AN APPLICATION OF GUTTMAN'S NEW SCALING TECHNIQUES	
TO AN ATTITUDE QUESTIONNAIRE	215
<i>Committee on Diagnostic Reading Tests, Inc. (with Frances Oralind</i>	
<i>Triggs)</i>	
DESCRIPTIONS OF THE PURPOSES AND FUNCTIONS OF THE	
DIAGNOSTIC READING TESTS	3
<i>Comrey, Andrew L. (with J. P. Guilford)</i>	
PREDICTION OF PROFICIENCY OF ADMINISTRATIVE PERSON-	
NEL FROM PERSONAL-HISTORY DATA	281
<i>Coopridge, H. A. (with H. R. Laslett)</i>	
PREDICTIVE VALUES OF THE STANFORD SCIENTIFIC AND THE	
ENGINEERING AND PHYSICAL SCIENCE APTITUDE TESTS	683
<i>Davis, Hilda</i>	
WHO SHOULD ATTEND OUR COLLEGES?	473

<i>DiVesta, Francis J. (with J. P. Hertel)</i>	
AN EVALUATION OF FIVE FACTORS FOR PREDICTING THE SUCCESS OF STUDENTS ENTERING THE NEW YORK STATE COLLEGE OF AGRICULTURE	389
<i>DiVesta, Francis J. (with Asahel D Woodruff)</i>	
THE RELATIONSHIP BETWEEN VALUES, CONCEPTS AND ATTITUDES	(645)
<i>Doppelt, Jerome E. (with George K Bennett)</i>	
THE EVALUATION OF PAIRS OF TESTS FOR GUIDANCE USE . . .	319
<i>Dudek, Frank J.</i>	
THE DEPENDENCE OF FACTORIAL COMPOSITION OF APTITUDE UPON POPULATION DIFFERENCES AMONG PILOT TRAINEES. I. THE ISOLATION OF FACTORS	613
<i>Durrance, Charles L.</i>	
MEASUREMENT IN RURAL HOUSING—A PROGRESS REPORT .	661
<i>Dyer, Henry S</i>	
SOME OBSERVATIONS ON THE COLLEGE BOARD LANGUAGE TESTS	593
<i>Edwards, Allen L.</i>	
ON GUTTMAN'S SCALE ANALYSIS	313
<i>Fruchter, Benjamin</i>	
THE NATURE OF VERBAL FLUENCY	33
<i>Guilford, J. P. (with Andrew L Comrey)</i>	
PREDICTION OF PROFICIENCY OF ADMINISTRATIVE PERSON- NEL FROM PERSONAL-HISTORY DATA	281
<i>Harris, Chester</i>	
A FACTOR ANALYSIS OF SELECTED SENATE ROLL CALLS, BOTH CONGRESS	583
<i>Hertel, J. P. (with Francis J. DiVesta)</i>	
AN EVALUATION OF FIVE FACTORS FOR PREDICTING THE suc- CESS OF STUDENTS ENTERING THE NEW YORK STATE COL- LEGE OF AGRICULTURE	389
<i>Heston, Joseph C.</i>	
A COMPARISON OF FOUR MASCULINITY-FEMININITY SCALES .	375
<i>Hopkins, E. H.</i>	
THE ESSENTIALS OF A STUDENT PERSONNEL PROGRAM .	430
<i>Humm, Doncaster G.</i>	
NOTE CONCERNING "THE VALIDITY OF STANDARD AND CUSTOM-BUILT PERSONALITY INVENTORIES IN A PILOT SELECTION PROGRAM," BY DONALD E. SUPER	(257)
<i>Hurd, A W.</i>	
IMPLICATIONS OF A BRIEF STUDY OF PREDICTION OF SUCCESS IN THE MEDICAL SCHOOL, MEDICAL COLLEGE OF VIR- GINIA	127
<i>Klein, George S.</i>	
SELF-APPRAISAL OF TEST PERFORMANCE AS A VOCATIONAL SELECTION DEVICE	(69)
<i>Kriedt, Philip H (with Kenneth E Clark)</i>	
AN APPLICATION OF GUTTMAN'S NEW SCALING TECHNIQUES TO AN ATTITUDE QUESTIONNAIRE	215

<i>Kvaraceus, W. C. (with Mary A. Lanigan)</i>	
PUPIL PERFORMANCE ON THE IOWA EVERY-PUPIL TESTS OF BASIC SKILLS ADMINISTERED AT HALF-YEAR INTERVALS IN THE JUNIOR HIGH SCHOOL	93
<i>Lanigan, Mary A. (with W. C. Kvaraceus)</i>	
PUPIL PERFORMANCE ON THE IOWA EVERY-PUPIL TESTS OF BASIC SKILLS ADMINISTERED AT HALF-YEAR INTERVALS IN THE JUNIOR HIGH SCHOOL	93
<i>Laslett, H. R. (with H. A. Coopridge)</i>	
PREDICTIVE VALUES OF THE STANFORD SCIENTIFIC AND THE ENGINEERING AND PHYSICAL SCIENCE APTITUDE TESTS	683
<i>Linton, Clarence</i>	
COUNSELING STUDENTS FROM OVERSEAS	501
<i>Mandell, Milton (with Sidney Adams)</i>	
SELECTION OF PHYSICAL SCIENTISTS	575
<i>Martin, Glenn C.</i>	
A FACTOR ANALYSIS OF THE BERNREUTER PERSONALITY INVENTORY.	85
<i>Miner, Robert J.</i>	
THE THERAPEUTIC HANDLING OF DISCIPLINE	550
<i>Norman, Ralph D.</i>	
A REVIEW OF SOME PROBLEMS RELATED TO THE MAIL QUESTIONNAIRE TECHNIQUE	235
<i>Portenier, Lillian G.</i>	
PREDICTING SUCCESS IN INTRODUCTORY PSYCHOLOGY	117
<i>Powell, Margaret G.</i>	
COMPARISONS OF SELF-RATING, PEER-RATINGS, AND EX- PERT'S RATINGS OF PERSONALITY ADJUSTMENT.	225
<i>Rimoldi, H. J. A.</i>	
A NOTE ON RAVEN'S PROGRESSIVE MATRICES TEST.	347
<i>Robinson, Francis P. (with Earl F. Carnes)</i>	
THE ROLE OF CLIENT TALK IN THE COUNSELING INTERVIEW	635
<i>Roeber, Edward C.</i>	
A MEANINGFUL RECORD OF TESTS	397
<i>Rogers, Carl R.</i>	
SOME IMPLICATIONS OF CLIENT-CENTERED COUNSELING FOR COLLEGE PERSONNEL WORK.	540
<i>Rosenberger, Homer T.</i>	
TESTING OCCUPATIONAL TRAINING AND EXPERIENCE	101
<i>Seeman, Julius</i>	
A STUDY OF CLIENT SELF-SELECTION OF TESTS IN Voca- TIONAL COUNSELING.	327
<i>Sherman, Jr., Arthur W.</i>	
PERSONALITY FACTORS IN THE PSYCHOLOGICAL WEANING OF COLLEGE WOMEN.	249
<i>Sims, Verner M.</i>	
QUESTIONING SOME ASSUMPTIONS UNDERLYING CURRENT ACHIEVEMENT TESTING.	565
<i>Sims, Verner M.</i>	
THE ESSAY EXAMINATION IS A PROJECTIVE TECHNIQUE	15
<i>Snyder, Harold E.</i>	
UNESCO AND EDUCATIONAL RECONSTRUCTION	496

<i>Sparling, Edward J.</i>	
MINORITY GROUPS IN OUR SCHOOLS	460
<i>Stone, C. Harold</i>	
ARE VOCATIONAL ORIENTATION COURSES WORTH THEIR SALT?	161
<i>Stromswold, Stanley A. (with C. Gilbert Wrenn)</i>	
COUNSELING STUDENTS TOWARD SCHOLASTIC ADJUSTMENT	(57)
<i>Tate, Merle W.</i>	
INDIVIDUAL DIFFERENCES IN SPEED OF RESPONSE IN MEN- TAL TEST MATERIALS OF VARYING DEGREES OF DIFFI- CULTY	353
<i>Taylor, Harold</i>	
HUMAN NATURE AND EDUCATION	530
<i>Traxler, Arthur E.</i>	
A PROCEDURE FOR OVERPRINTING ANSWER SHEETS FOR HAND SCORING WHICH MIGHT BE ADAPTED TO LOCAL SCORING	65
<i>Triggs, Frances Oralind (with Committee on Diagnostic Reading Tests, Inc.)</i>	
DESCRIPTIONS OF THE PURPOSES AND FUNCTIONS OF THE DIAGNOSTIC READING TESTS	3
<i>Warren, Neil D. (with A A Canfield)</i>	
AN OPTOMETRIC APTITUDE TEST	183
<i>Welsh, William B.</i>	
DEMOCRACY THROUGH STUDENT GOVERNMENT	522
<i>Williamson, E. G.</i>	
SUPERVISION OF COUNSELING SERVICES	297
<i>Williamson, E. G.</i>	
THE GROUP ORIGINS OF STUDENT LEADERS	603
<i>Woodruff, Asahel D. (with Francis J. DiVesta)</i>	
THE RELATIONSHIP BETWEEN VALUES, CONCEPTS AND ATTITUDES	645
<i>Wrenn, C. Gilbert (with Stanley A. Stromswold)</i>	
COUNSELING STUDENTS TOWARD SCHOLASTIC ADJUSTMENT. .	57
<i>Wrenn, C Gilbert</i>	
THE GREATEST TRAGEDY IN COLLEGE PERSONNEL WORK . . .	412

EDUCATIONAL and
PSYCHOLOGICAL



MEASUREMENT

VOLUME EIGHT, NUMBER ONE, SPRING

<i>Description of the Purposes and Functions of the Diagnostic Reading Tests</i> COMMITTEE ON DIAGNOSTIC READING TESTS, INC.	3
<i>The Essay Examination is a Projective Technique.</i> VERNER M. SIMS	15
<i>The Nature of Verbal Fluency.</i> BENJAMIN FRUCHTER . . .	33
<i>Practical Suggestions for Improving Testing Procedures.</i> LEO CHALFEN	49
<i>Counseling Students toward Scholastic Adjustment.</i> STANLEY A. STROMSWOLD AND C. GILBERT WRENN	57
<i>A Procedure for Overprinting Answer Sheets for Hand Scoring which Might be Adapted to Local Scoring.</i> ARTHUR E. TRAXLER	65
<i>Self-Appraisal of Test Performance as a Vocational Selection Device.</i> GEORGE S. KLEIN	69
<i>A Factor Analysis of the Bernreuter Personality Inventory.</i> GLENN C. MARTIN	85
<i>Pupil Performance on the Iowa Every-Pupil Tests of Basic Skills Administered at Half-Year Intervals in the Junior High School.</i> W. C. KVARACEUS AND MARY A. LANIGAN ..	93
<i>Testing Occupational Training and Experience.</i> HOMER T. ROSENBERGER	101
<i>Predicting Success in Introductory Psychology.</i> LILLIAN G. PORTENIER	117
<i>Implications of a Brief Study of Prediction of Success in the Medical School, Medical College of Virginia.</i> A. W. HURD .	127
<i>New Tests</i>	133
<i>Measurements Abstracts</i>	145
<i>The Contributors</i>	155

COMPLETE BACK VOLUMES
NOW AVAILABLE!

In response to a considerable demand for out-of-print copies of EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, Volumes I through IV have been reprinted in a small-print edition which is available at \$3.00 per volume (paper-bound). The regular page size has been retained, so that these volumes match the regular volumes except in thickness. Four pages of the original have been reduced to one page of the small-print edition. The reduced print is readable, and will be found to be quite adequate for most research purposes. A sample page will be sent on request.

Volumes V, VI and VII are available in the regular edition at \$5.00

Please send orders to EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 917 Fifteenth Street, Washington 5, D. C.

DESCRIPTION OF THE PURPOSES AND FUNCTIONS OF THE DIAGNOSTIC READING TESTS

FRANCES ORALIND TRIGGS, CHAIRMAN¹

THE purpose of this paper is to report the construction of a battery of diagnostic reading tests and to indicate areas of projected research.

The work of the Committee on Diagnostic Reading Tests, the body of specialists responsible for the construction of these tests, was recently described in some detail in an article in *School and Society*.² It will be given here only to the extent that it affects the present report.

The purpose of the Committee in constructing these tests was to provide tools for surveying the reading skills of students from the seventh grade through the college freshman year in order that (1) a diagnosis of any existing disabilities may be made and appropriate remedial instruction instituted for retarded readers and that (2) reading instruction and materials may also be better adapted to the skills of generally normal pupils.

The battery of tests was developed by a committee of specialists in the field of reading under a grant of money from the Blue Hill Foundation. This grant covered the development of materials only—not their publication. The proceeds from the sale of materials now available will be used to finance the publication of the remaining tests of the battery.

¹ This article was prepared by the Committee on Diagnostic Reading Tests, Inc., which includes the following members

Robert M. Bear, Dartmouth College
Ivan A. Booker, National Education Association
Daniel D. Feder, University of Denver
Constance M. McCullough, San Francisco State College
A. Eason Monroe, San Francisco State College
George D. Spache, Chappaqua, New York Schools
Arthur E. Traxler, Educational Records Bureau
Frances Oralind Triggs, Educational Records Bureau, *Chairman*

² Triggs, Frances Oralind "Diagnostic Reading Tests as Aids to Remedial Instruction," *School and Society*, LXVI (1947), 42-45.

The plan of the Committee is to provide for continuous research on the tests and for revision of them as needed. Since the project is entirely a nonprofit one, all proceeds from the sales of the tests in excess of publication costs are to be devoted to these purposes.

The members of the Committee on Diagnostic Reading Tests are: Robert M. Bear, Dartmouth College; Ivan A. Booker, National Education Association; Daniel D. Feder, University of Denver; Constance M. McCullough, San Francisco State College; A. Eason Monroe, San Francisco State College; George D. Spache, Rohrer, Hibler and Replogle, New York; Arthur E. Traxler, Educational Records Bureau; and Frances Oralind Triggs, Educational Records Bureau, Chairman.

All sections of this battery are adapted for both machine and handscoring except the oral test. Tentative percentile and grade norms are provided, but it is hoped that schools using the tests will send scores to the Committee in order that these norms may be supplemented and more extensive data for further research may be available.

The following outline indicates areas which the battery of tests cover:

1. *Survey Section*—A survey or screening test to provide a reliable instrument for determining the general level of reading achievement of pupils in junior and senior high school and in the freshman year of college, including measures of rate of reading simple material of general interest, meaning vocabulary, and comprehension of textbook material.

2. *Diagnostic Sections*—A series of diagnostic tests, to be given when reading disabilities are discovered through the use of the survey test. The diagnostic tests cover the following areas: vocabulary, auditory and silent comprehension, rates of reading under varying conditions and of various types of subject matter, word attack and word recognition skills.

Section I of the diagnostic test battery, the *Vocabulary Test*, yields scores in the following five areas: general vocabulary, vocabulary of English grammar and literature, vocabulary of mathematics, vocabulary of science, and vocabulary of social studies,

Section II, *Comprehension*, is presented in two parts, Silent and Auditory. The tests measure comprehension of textbook-type material drawn from the social studies, science, and literary fields. The reading material is graded in difficulty according to the Lorge formula and according to more subjective criteria such as composition, structure, and type of content. From scores on the two parts of the test, Silent and Auditory, a comparison may be made between the pupil's understanding when he reads material silently and his understanding when it is read to him. Thus, some estimate can be made of the extent to which a reading deficiency may be handicapping a pupil's ability to comprehend what he reads, and the extent to which insufficient maturity and poor background make it inadvisable to attempt to present ordinary textbook materials.

Section III of the tests, *Rates of Reading*, has three parts. Part 1 measures the pupil's usual rate of reading interesting, story-type material, as compared with his ability to adopt a faster rate of reading and still comprehend comparable material when he is instructed to vary his rate according to a different purpose. On both parts of this test a check is made on the pupil's comprehension of the material read. Part 2 measures the rate at which the pupil usually reads and comprehends social-studies material. Part 3 measures the pupil's usual rate of reading with comprehension, science material. These areas correspond to three of the areas of vocabulary measurement in Section I.

Section IV of the tests, entitled *Word Attack*, measures word-recognition and attack skills and has two parts, Oral and Silent. Part 1, the Oral Test, is the only individually administered portion of the battery. Six graded paragraphs of interesting, general-type reading material are read aloud by the student. The teacher observes the pupil's reading attitude and methods and marks the errors he makes while reading, using a modified method of scoring based upon that of the *Gray Oral Reading Paragraphs*.

Section IV, *Word Attack*, Part 2, Silent, attempts a new approach to the measurement of word recognition. This part

of the test is set up for group administration. Through it an attempt has been made to measure the ability to hear sound by using a matching technique and to test the pupil's ability to divide words into syllables. Remedial techniques along these lines can be used profitably to improve the reading skills and also the spelling proficiency of some pupils, and remedial exercises are already available which classroom teachers can use successfully.⁸ It is not yet known how such skills are related to the total reading process, but this is being studied. The instrument we have built is, however, a reliable one.

From this battery of diagnostic tests it should be possible for an instructor to select tests for use with the pupils whose reading development is unsatisfactory, to determine what skills are inadequate and why the development of their reading has not kept pace with what would be ordinarily expected. With the aid of such diagnostic tests, and other information concerning the individual, classroom instructors should be able to undertake remedial instruction, with at least some of their own pupils.

There are a number of areas in which diagnostic tests *might* have been set up. However, the Committee, in studying the problem, had in mind that the results of the tests should provide immediate clues to remedial instruction. Therefore, in building the tests, initial attention was given to those basic areas where remedial work has already proven to be successful. However, they did not include tests measuring specific skills such as graph-reading because this skill, while important to efficient, critical reading in the higher grades, can usually be taught in a session or two of group work after good basic reading skills have been developed. Unless a student has the essential basic reading skills, the ability to read graphs will not greatly profit him. Therefore, they did not consider measuring such specific skills the most efficient use of testing time, though a teacher might undertake it as a classroom technique in group reading instruction.

The Committee has been asked why the tests were not constructed along lines seemingly indicated by factor analyses of reading skills. The answer is that while in the future the

⁸ Triggs, Frances Orland *Improve Your Spelling*. New York Farrar and Rinehart, 1944.

Committee is very much interested in any research which will advance the measurement of reading, its main immediate concern is to aid in reducing the lag between what is now known concerning the teaching of reading and what is done in the secondary grades and college. From this point of view, it seems better to measure the groups of skills basic to developmental reading which we know the personnel worker and teacher can do something about than to measure pure factors (supposing we found we could do so practically in the average class room) and then not know how to approach profitable instruction along the lines the measurement dictated. Which approach would bring the desired results most quickly may be debatable, but the reasoning of this Committee early led it to discard the factor analysis approach. It is undoubtedly true, however, that members of the Committee expect and will rejoice if, through the approach they have taken, dictated by practical consideration of immediate needs, comes a better scientific understanding of the reading process and ways of measuring pure factors basic to reading, and perhaps a better understanding of their development through the educational process.

All tests constructed for the preliminary experimental edition were subjected to the usual item validation techniques and reliabilities were checked. All tests were given to the same students in order that intercorrelations of all parts might be run. These data were then studied and a selection of material for the final experimental edition was made. This edition was again subjected to analysis using the same item validation techniques. Reliabilities were again run and intercorrelations with a general measure of ability were computed. From these data available on the second experimental edition, the final edition of the tests was constructed.

The *Survey Section*, and Section IV, *Word Attack*, Part 1, Oral, and Part 2, Silent, went on sale September 1 and were distributed by the Educational Records Bureau, a nonprofit educational organization, in New York. All other sections of the diagnostic battery will be available June 1, 1948, from the same source.

The Committee is now incorporated as a nonprofit membership corporation under the laws of the State of New York for the purpose of research in testing and the use of test results in remedial and developmental reading programs, and in aiding schools to better utilize the results of such research, tests and remedial aids as are available to them.

The remainder of this paper indicates some of the research which has been done and the research now pending. This outline of projected research will, of course, be expanded and modified as the work develops.

The *Survey Test* is intended as a screening test to be used with all students and was built to measure a composite of the most usual types of reading skills necessary for use in school work. Certain types of data are now available on this test which are indicative of its validity for use as a screening test.

The vocabulary and subject matter used in the test, are of the nature which students regularly meet with in school. Comprehension questions were constructed with a view to measuring main ideas, details, and reading of a critical nature.

From an administrative point of view, a test of this nature needs to be so constructed that testing time is efficiently used. Average validity coefficients of the 100 items in the test are: Form A, 54; and Form B, 53. Average percentage of difficulty of the items for ninth-grade students is about fifty-four.

The *Survey Tests* were given to approximately 3,000 students in seventeen schools of wide geographical distribution to establish tentative percentile and grade norms. Reliabilities estimated by Kuder-Richardson Formula 21 at each grade level indicate that the total comprehension score is a reliable measure. The average of coefficients obtained is .90. The average of coefficients obtained for rate of reading scores is .80.

Part-score norms are being furnished on this section of the test in order to provide clues for judging which sections of the diagnostic battery should be administered. However, because it was thought important from an administrative point of view

to restrict testing time to one class hour on the *Survey Test*, it was not possible to include enough items of each type, comprehension check on rate of reading, vocabulary, and comprehension of textbook-type material, to make these scores completely satisfactory for diagnosis in individual cases. The directions for administering and interpreting this test warn against using them for that purpose.

Plans for research include the investigation of certain relationships, suggestive of validity, such as:

1. Relationship to verbal and quantitative scores on a well-established test of general ability.

2. Extent of progress made from test to retest when *usual type* of instruction in reading is given over specific periods of time.

3. Extent of progress made from test to retest when well-defined types of *remedial instruction* are given.

4. Relationships to other well-established survey tests.

The diagnostic sections of this battery were set up after making a survey of the types of test data specialists in remedial reading felt they needed on which to base remedial instruction and after surveying the areas in which it has been found that successful remedial instruction can be done.

Section I of the diagnostic battery is a vocabulary test made up of five parts, each yielding a reliable score: general vocabulary, vocabulary of literature and English, of mathematics, of physical sciences, and of social sciences. The area in which items were placed was determined by use of word lists. Each item presents the key word in context. The correct response is a synonym of the word as *used in the context* in which it is presented. The test is untimed in order that scores on it will be indicative of the maximum ability of the student. The subject areas in which vocabulary is tested correspond (except for mathematics) to the areas in which measurement is made in comprehension and rate of reading.

Items for Section I, Vocabulary, have been validated, but since norms are not set up, not much statistical data can be reported at this time. On the preliminary form intercorrelations between part scores were high. Correlations with a meas-

ure of general ability (the *Terman-McNemar Group Test*) approximately .68. Correlations with word recognition as measured both orally and silently were generally low.

This section of the test will be useful with students deficient in the technical vocabulary needed for efficient reading in the several areas measured. While it is known that it is very hard to show improvement on vocabulary test scores through remedial work in general vocabulary, aid in specific areas does seem to bring both measured improvement and helps to establish techniques for obtaining meanings from strange words. This section of the test is in final form. Norms are being set up and the *Vocabulary Test* will be ready to go on sale in June, 1948.

A great deal of work has been done on the Comprehension section of the diagnostic battery. Those who have worked in this area know the difficulties of constructing tests in this area. Comprehension of material read, as usually tested, is closely related to general ability as measured by our so-called scholastic aptitude or general ability tests. Both types of measures are closely related to cultural background. The Committee is aware that the measurement of ability to comprehend as against actual comprehension should be distinguished because potentiality for comprehension is an important factor in remedial work. How can it be determined, using paper and pencil techniques, to what extent these measures are affected by these two factors? In this battery, an attempt has been made to differentiate the factors by varying the method by which the material is presented. Comparable material is read by the students and is read to them. It is not known, however, to what extent listening is a learned skill. Since listening to material being read aloud is not the usual technique used by students in preparing textbook assignments, such additional factors as concentration, skill in attack on new words, etc., may result in these two approaches measuring somewhat differing processes. Much research will have to be done along this line, even after the technical difficulties of constructing this section are surmounted.

It was also requested by persons doing remedial work that the *Comprehension Test* give reliable and valid subscores on

such factors as ability to obtain the main thought of a passage, ability to support the main thought by details as presented by the author, ability to distinguish between important and unimportant details as given. Some would also like to be able to distinguish these skills as applied to clearly defined types of content such as social science, science, and literary material. As yet, our finding is that it has not been possible to distinguish between such factors when measured—in other words, correlations of scores on items of good validities are so high as to make it questionable whether any distinction is, in fact, being made.

Section III of this battery measures, to the extent that the instructions actually cause students to respond in the manner indicated, the usual rate of reading with understanding material of a general story-type, material utilizing social science vocabulary and concepts, and material utilizing physical science vocabulary and concepts. Data from this section, along with data from other sections of the diagnostic battery, especially vocabulary, give important clues for remedial work.

The first part of Section III does more than measure the "usual" rate at which students recall and comprehend the material presented. That is only the first task presented to the students. After their "usual" rate of reading is measured on a passage using material of a general story-type, they are presented with material comparable to that of the first passage with instructions to read it as fast as they possibly can and comprehend it. Thus, by comparing the scores a student makes on the first and second passages, it should be possible to get some clue as to the extent to which he is flexible in applying his skills, a very important factor in efficient, all-around reading. Is the technique adequate to our task? The answer is not yet known. Only further carefully controlled research will tell.

Another question has been raised. Ability to skim as a reading skill makes for flexibility in reading. To what extent does this latter approach measure skimming as against more thorough reading? Perhaps the extent of comprehension, given equal previous knowledge about both passages, will help to answer this question.

Another possible approach to controlling the actual reading process more closely would be to set the stage by raising certain questions before the passages were read, the answers to which students would read for. The process might even be more closely controlled by presenting the student with questions on the topic covered by the passages before he reads them, for perhaps we are assuming his comprehension comes from reading, when actually it comes from previous knowledge. Such questions might be presented some days before the actual "reading" test was administered if presentation of the questions was thought to affect the reading process in a manner not consonant with the purpose of the test.

The literature is not by any means devoid of research reports on the relationships existing between the various factors related to the rate at which material is read and other measured reading skills. However, all of the factors measured have not always been carefully controlled. Therefore, much can be added to our knowledge in this area. It is expected that tests in this Section will also be ready for sale by June, 1948.

The last general area in which tests are now set up in the diagnostic battery includes word attack and word recognition. It is perhaps justifiable to say that basic to all reading is the ability to recognize word symbols previously learned and the skill in attacking words not so recognized. There are a number of ways in which a thought can be obtained. No person can be an efficient reader unless he has some skills for attaining thought from unknown words. He must know how to obtain meaning from those whose configuration alone give no clues. It is these skills which the *Word Attack*, Section IV, of this battery of tests measures.

It has long been agreed that a significant value of oral reading is to test these basic reading skills and the pupil's methods and attitudes of approaching reading. Through analysis of errors of mispronunciation, omission, confusion of words of like configuration and such factors, the reasons for inaccurate comprehension may become apparent when otherwise only the fact that comprehension was inaccurate may be the only thing recognized. Therefore, the battery provides an oral reading

test of six graded paragraphs carefully controlled from the point of view of interest, vocabulary load, length, etc. These are presented orally to the testee and errors of various types are recorded according to a somewhat usual standardized practice.

A great deal of research needs to be done with the data that such a test will provide. For instance, we do not yet know whether secondary-school grade norms on this part of the test are needed, for present evidence suggests that there may be as much variability within grades as between grades and that there is not much change in these basic skills at successively higher grade levels. It was not known to what extent these types of errors could be reliably measured and therefore their use justified in differential diagnosis and remedial work, until results were in from the experimental edition which showed reliability of scores in approximately 100 Cleveland ninth-grade pupils of over .90.⁴

The question also arises as to whether the approach a student makes to words known or unknown to him might better be revealed by presenting to him words in list form rather than context, for context does present an uncontrolled variable. Therefore, at present our word attack test also provides three word lists to be pronounced. Research will be done in order to determine the extent to which either one or both measures are getting at what the test is purporting to measure.

It seems possible to measure certain word recognition skills by group techniques rather than by oral, individual presentation, thus perhaps controlling the situation more closely and also utilizing testing time more efficiently. Therefore, Part 2 of the *Word Attack* Section, utilizing a matching technique, attempts to measure the ability of the student to hear sound, it being concluded that ability to sound words is one factor in approaching a strange word. This part of the test also measures the ability of students to divide words into syllables.

The relationship which these last two skills have to related skills as measured by oral reading and by the skills measured by other sections of the test is almost unscratched from the

⁴ By checking errors in odd vs. even lines and using the Spearman-Brown formula.

point of view of research. We do know that students who cannot divide words into syllables or who cannot hear sound, or both, when taught these skills make immediate and surprising progress in many phases of reading. But this is a non-statistical validating technique. From a statistical approach about all we know is that this measure as constructed is a highly reliable one and that correlations between it and other tests in the battery when in experimental form ranged from zero to .60, with other sections of the *Word Attack* test, and about .50 with vocabulary and other sections of the battery. The range of correlations with outside criteria such as a group test of general ability and vocabulary, comprehension as measured by scores in another survey type reading test, was about the same. It is therefore apparent that there is as much room for fruitful research to be done on this part of the test as upon the others.

This battery of tests affords a great many possibilities for a more exact definition and measurement of reading skills and for measurement of growth in reading. Such instruments are needed if reading is to serve efficiently as a tool in the greatest task we face, that of helping each member of societies the world around to make his contribution to a lasting peace.

•

THE ESSAY EXAMINATION IS A PROJECTIVE TECHNIQUE

VERNER M. SIMS
University of Alabama

IN spite of more than three decades of criticism (one might almost say abuse) by educational measurers, the essay examination continues today strong and healthy. In the face of a long contention that it does but poorly what the short-answer or so-called "objective" test will do better, even today probably more of the achievement testing used by teachers, at least at the upper levels of the educational system, is of the essay rather than of the "objective" type. Such vitality certainly suggests some uniqueness of function on the part of the essay examination, however vaguely and inadequately this function may have been identified. And, by and large, it seems safe to say that this type of examination has been used uncritically, without any very clear realization of its functions and without much effort to improve its measuring qualities. This paper represents an attempt to identify and to define more clearly what appears to be a unique function of the essay examination, to the end that later research and investigation may lift its use to a more critical and therefore more helpful level.

Although the fact has not yet been noted in the literature on educational measurements, it seems clear that in a scheme of human measurements, the essay examination would properly be classified as a "projective" method of measurement. Since projective methods, as such, are relatively unknown in the field of achievement testing, some justification of this statement appears to be in order. Let us begin by trying to identify projective techniques.

Frank, who seems to have introduced the word *projective method* into the psychological literature, defines the term as follows:

A projective method for the study of personality involves the presentation of a stimulus situation designed or chosen because it will mean to the subject not what the experimenter has arbitrarily decided it should mean (as in most psychological experiments using standardized stimuli in order to be objective) but rather whatever it must mean to the personality who gives it, or imposes upon it, his private, idiosyncratic meaning and organization (3, p. 403).

As the name implies, projective tests involve problematic situations which permit (almost require) the testee to "project" his personality into his answer. Such methods are commonly contrasted with "objective" or short-answer testing. The media in use for projective testing may consist of such organized materials as personal documents, dramatic or artistic forms, and constructions; or they may consist of relatively "unstructured" materials such as ink blots, finger paintings, and even the free use of cold cream.

The test situations are characterized by a considerable "freedom of choice" on the part of the testee, and by the relatively complex nature of the response demanded. Rather than pre-determined "right" responses which must be recognized or recalled in order to do well, in projective testing the testee is compelled to make choices in terms of his own experiences and sense of values.

The concern of the tester is typically with the "pattern" of response made rather than with a single summation of "right" choices, hence information concerning the "why" and the "path" of choice is available. Both the "manifest content" made intentionally in response to the problematic situation, and the "hidden content," reflecting expressive and stylistic performance and unintentionally revealing aspects of the mental life, are considered legitimate data from which to draw inferences concerning the personality of the testee. Thus inferences concerning "characteristic ways of behaving" or psychological "dispositions" are drawn from both classes of data. In fact, many users of projective methods consider the "hidden" con-

tent as the more valuable source of insight. Finally, projective testing is done more commonly for the understanding which it furnishes of the individual personality than for "normative" purposes.

The reader who cares to try it will be amazed at how neatly the concept of the essay examination can be inserted into the above characterization of projective methods. The best definition of this type of test which the writer has been able to produce is almost a paraphrase of Allport's definition of the personal document (1, p. 12), which he recognizes as a type of projective method:

The essay examination is a relatively free and extended written response to a problematic situation or situations (question or questions), which intentionally or unintentionally reveals information regarding the structure, dynamics and functioning of the student's mental life as it has been modified by a particular set of learning experiences.

The necessity to bring to bear the "higher-order" aspects of one's learning and the need of a relatively complex and, therefore, long answer in order to do this has been fairly well recognized and defended by users of the essay question. Clearly the essence of the difference between essay and objective tests is in terms of the freedom and extended nature of the response in essay testing. The unintentional revelations which essay answers have for the "initiated," although not too consciously recognized, have also been an important factor in the continued popularity of the instrument.

One element in the definition given does seem to assure the legitimacy of the essay examination as a test of educational achievement and, at the same time, to point out its uniqueness as a projective technique. This element, contained in the last phrase, is the concern over the "structure, dynamics and functioning of the student's mental life as it has been modified by a particular set of learning experiences." This is not the purpose for which projective testing has been used; but it is the end which all educational measurements commonly attempt to serve. And incidentally, it is an end which many teachers feel is not too completely served by the atomistic and oftentimes

superficial measurement which results from short-answer testing.

To identify the essay examination as a projective method may perhaps but strengthen the conviction of many educational measurers that the instrument occupies an unhappy place in any scheme of psychological testing. On the other hand, to so classify it seems to define more clearly those characteristics of the instrument which have been involved in its defense but not recognized too well by its defenders. Moreover there is a growing conviction on the part of many that projective testing has an important part to play in psychological science, a conviction that the critical and intelligent use of projective methods furnishes insights which are not available (and by the nature of things cannot be available) in the data furnished by restricted responses to "objective" questions.

One has but to examine the recent literature to appreciate the place that projective testing has come to take in the researches on psychological measurement. Sargent, for example, in a recent review of literature on the subject lists a bibliography of 274 titles most of which have been published in the last few years (4). It is important also to note that a number of investigators have subjected projective methods themselves to careful study, with the result that, to a degree at least, a critical and intelligent use of them is now possible.

If it is accepted that projective methods do have a useful place in human measurements, identifying the essay examination as such a method makes it immediately possible to suggest applications of the findings concerning the critical use of such instruments already at hand from the general field of projective measurement. Ultimately, it should lead to the development of essay testing as an important and legitimate part of the general field of projective testing.

The remainder of this paper represents a first attempt at the former of these tasks. In suggesting procedures for using the essay examination we shall lean heavily on Allport's *The Use of Personal Documents in Psychological Science* (1). The truth is, the writer's "insight" relating to this problem came chiefly as a result of reading Allport's monograph which might

well become "required reading" for students in the field of educational measurement.

The reader will appreciate the fact that the suggestions offered here are to be accepted as tentative. That is, they are generalizations from the field of projective measurement which are as yet relatively untried in their application to the specific field of essay testing. They do, however, appear as reasonable hypotheses worthy of being tested in practice. The suggestions fall naturally into three classes: (a) those relating to the kinds of outcomes to test for; (b) those relating to the nature of the test questions; and (c) those relating to the handling of test data.

I. Suggestions Relating to the Kinds of Outcomes for the Measurement of Which the Essay Examination Is Suited

Indiscriminate and uncritical use of the essay examination, preparing questions without carefully identifying the learning outcomes concerning which evidence is being sought and without thought as to the suitability of the essay-type question for securing such evidence, has undoubtedly been common. All too often the accusation that essay testing is the lazy man's way of doing it seems justified. Intelligent use of the essay examination demands that we first isolate the particular outcomes of learning for which we wish to test and then that we inquire as to the appropriateness of this instrument, along with others, for doing the job. The following suggestions should help with this task:

1. The essay examination appears to be particularly well suited for obtaining evidence relating to certain "higher-order" intellectual outcomes of education.

Although most essay testing actually done is probably primarily concerned with the recall of information learned, the value of the essay for testing ability to organize, relate, and "weigh" materials learned has been long appreciated. These, however, are but a few of a number of higher mental processes for which the test seems well adapted. To name others, where the concern is over the extent of which particular learnings are

integrated with previous learning, over the "distance" that particular learning can be transferred, or with the ingenuity of response (the ability to use learning creatively), the essay examination seems called for. Although clever recent developments in short answer testing suggest that some of these outcomes may be susceptible to "objective" measurement, to date it seems to be chiefly to the essay that we must turn for information in these fields. In cases where information on the "path" of solution followed voluntarily and without suggestion is desired, such as is involved in the "elegance" of solutions in mathematics, the free situation provided in the essay examination seems required. And, finally, when we want to know the "frame of reference" with which a student approaches problems (information which becomes highly important when we undertake to predict whether he will *use* his learnings), the freedom of the essay is demanded. Objective tests, with their predetermined "right" responses predetermine also the frame of reference with which the student must approach the problem. Contrast, for example, the objective question beginning, "What scientific principle is involved in the following situation?" with an essay beginning, "What would *you* do in the following situation?" Nor, if we want information on whether a student *will* use his learning, do we get it by converting the above essay question into an objective test with alternate choices! What we want to know is how he approaches the problem *voluntarily and without suggestion*.

2. Inferences concerning certain personal-social learnings are perhaps best drawn from the kind of data obtained through essay questions.

In a sense one might say that the usefulness of the essay examination varies directly with the breadth of the responsibility assumed by the teacher. If the end sought in teaching is changed behavior on the part of the learner, then the fundamental question in all educational testing is: Can we, in terms of the evidence at hand, predict that the learner will act differently in a certain class of situations? Once the matter is stated in this manner it becomes evident that adequate testing will always comprehend more than *ability* to do. Whether one *will*

act so and so depends not alone on whether he can do so but also on a complex of attitudinal-emotional-motivational factors which may be loosely called personal-social in nature. A comprehensive testing program will include data from which inferences concerning these learnings may be drawn. It is here, perhaps, where essay examinations, along with other projective methods of measuring the outcomes of learning, have their greatest usefulness. The freedom and complexity of the essay typically *require* the testee to bring his own "sense of values" to bear on the task-provoking situation. In the well-planned question the respondent not only *does* but *must* reveal information which is pertinent to any predictions concerning how, when, and where the abilities he has acquired will be used.

Furthermore, many teaching situations today are consciously planned not as means for intellectual development but for the purpose of modifying attitudes or rebuilding emotional patterns. In this teaching curriculum experiences are built directly and immediately out of the purposes and goals of the individual learner. In such situations the essay in some of its manifestations seems to be the only "paper-and-pencil" testing technique available. The varied and highly personal nature of such learning can only be revealed through a "personalized" test situation, and even then oftentimes only through the "hidden" content.

3. By and large the essay examination would seem to be more useful for the information which it reveals concerning the individual than for "normative" purposes such as are commonly involved in "grading."

Where "grades" are interpreted as expressions of relative position in a class, essay examinations can probably be so used as to be reasonably dependable. In fact, it has been proposed that the basis for making a decision concerning whether to use objective or essay tests should be economy of time for the teacher—in small classes the essay would be the more economical; in large classes, the objective. In general, however, this writer is not convinced that the particular forte of the essay examination is for purposes of grading or as a final examination. Instead, its strength seems to be in terms of its diagnostic val-

ues. It is in terms of the information revealed concerning particular individuals, information which will be helpful in working further with these individuals. As commonly used, objective tests have a minimum of diagnostic value. Granted that some improvement can result from a more intelligent use of them, they still have one serious limitation—the information revealed can never go beyond that which is anticipated by the test-maker. On the other hand, the unique value of the essay is often in terms of the unexpected insights revealed; revealed by the naive subject in both the manifest and hidden content, but revealed, to the initiated at least, in the hidden content of the responses of the most sophisticated students.

As a means of illustrating this point let the reader consider what might result from developing a series of doctoral "preliminaries" which were deliberately and systematically designed to get at the intellectual and personal-social learnings of the student; to the end that we might get leads to further work with him and for him, and with the matter of "passing" or "failing" being purely incidental. Suppose, for instance, that the readers of preliminary examinations were required to prepare a list of suggestions for the further professional development of the student, suggestions for him and for those who have responsibility for his training. From objective tests, if we had enough cases to make our interpretations dependable, we might get information on what he *does* or *does not* know and *can* or *cannot* do. We might be able to say that he needs to study more in this or that field; or, if we have norms for part scores (and this means a still larger number of cases), we might go so far as to say that his study should be on this subject or, if our breakdown was small enough, on this topic. Perhaps, even, if we used some of the more recently developed techniques, we could say that he needed to acquire understanding, to learn to apply principles, to reason logically, or to interpret data, in this field or that field. At best, the information revealed through objective tests would be limited to the student's *abilities*. This information is in sharp contrast to the richness of the insights concerning his motivations, his attitudinal patterns, and his habits of action which could come from well-prepared exhaustive essay examinations intelligently handled.

And it is to the credit of the essay examination that we cannot predict precisely the nature of these insights prior to an examination of the papers. What we can do is to identify some of the intellectual and personal-social characteristics which we consider important, to plan a variety of free-response problematic situations (questions) which appear fruitful for eliciting information relating to these characteristics, and then to develop systematic means of handling the test data so as to make usable whatever information inheres in them. And it is such procedures which need so badly to be subjected to experimental try-out by our testing experts.

4 Finally, it would appear uneconomical, in terms of time and energy and in terms of the level of skill demanded in handling the answers, to use essay tests for securing information which can be obtained through short-answer testing

When large numbers of students are involved, the time and labor necessary for preparing, administering, scoring and interpreting objective tests is obviously less than that needed for essay tests. Furthermore, the level of skill required for a good part of the work in handling objective tests (that of scoring) is of a low order and can be delegated to clerks or other assistants. These facts mean that whenever objective tests will do the job, as, for example, in measuring factual learning, they are to be preferred. They mean, also, that the tester's skill and ingenuity in objective testing and his facilities for such testing become a factor in determining the extent to which essay tests will be used. They do not mean, however, that we should select for testing only those outcomes which can be tested by short-answer methods. It may be that for crude classification purposes, such as are involved in a 5-point grading scale, for example, the correlation among desirable outcomes (which are known to exist) are sufficiently high to justify one in sampling only those learnings which can be tested objectively. Even this appears risky in the face of evidence which is at hand concerning the influence that the type of test used by the teacher has on what students study and how they do it. Certainly when we want to understand the individual and the nature of his learning, the economy of short-answer tests is no

justification for substituting evidence relating to one outcome for evidence relating to another which can be more easily tested!

II. *Suggestions Relating to the Nature of Good Test Questions*

The framing of "good" essay questions, (that is, questions which do cause the testee to reveal, both intentionally and unintentionally, the nature of his mental life as it has been modified by certain learning experiences) is obviously one of the most important jobs in essay testing, and one for which really dependable rules will have to await further investigation and study. However, certain suggestions are proposed:

1. Essay questions should permit a relatively free response and should encourage an extended response.

In objective testing, it is the rule to limit rigidly the requirements of the adaptive task (question) to the end that there be only one "right" response. If essay questions are to be most useful as projective instruments, exactly the opposite rule should hold. As Allport puts it, "The more exacting the prescription in terms of task, the less value the instrument has in terms of expression or projection." (1, p. 112.) If we want to find out what learning a student will bring to bear on a question which is posed to him, then he must have freedom of choice, and the greater the freedom the more chance that valuable insights into his learning will be revealed by his answer. Contrast, for example, these two questions: "Discuss the Articles of Confederation," and, "Discuss the Articles of Confederation with respect to their origin, their working out in practice, and their relationship to the present federal constitution." The second question is obviously designed to elicit specific learnings of a factual nature relative to the Articles; but if one is interested in getting evidence relating to the extent to which the student has integrated his readings dealing with the Articles with other learnings, or evidence bearing on his attitudes, or on his ability to personally evaluate the Articles, then the chances are better with the first question. This because of the freedom with which it permits him to project his own personality into the situation.

Furthermore, within the limits of available time (for taking and for reading), the more the student writes, the better. The relation between short answers and objectivity is often assumed to be so intimate that the terms "short answer" and "objective" have become more or less synonymous in testing. But if essay tests are to be most valuable as projective instruments we must encourage the student to write *at length*. If we give him enough rope he may hang himself, but he may, too, reveal just the information which we need and cannot get through our objective testing. We will have to seek agreement among competent judges not by arbitrarily limiting the response but, instead, by agreement as to the outcomes concerning which we are seeking information and by agreement as to how we will handle the test data. That is the only real sense in which we have objectivity in educational measurements anyway.

2. Generally, essay questions should pose problems, and problems that have a "reasonable" separation from the original learning situation.

Essay questions which call for the uncritical recall of facts learned or questions for which the students have been directly coached (types of questions which are too common) seem scarcely worth mentioning. But aside from these, some essay questions put the student in a new situation and one to which learning acquired in a different setting must be brought to bear, while other questions demand that he respond directly in an interpretative or evaluative manner to some particular learning experience. Illustrative of the first would be "problems" in mathematics, new situations involving principles learned in science, new situations in which attitudes may (or may not) color the response, situations demanding critical choice from total materials learned, etc. The second type would be exemplified by questions beginning with such phrases as: "Tell the story of," "What is your opinion of," "What do you think of," "Tell what you know about." Psychologically these questions are "free-responses," and as such have their value; but they furnish only limited (and indirect) evidence concerning the extent to which a student will transfer his learning. Generally speaking, it would seem better to use the "new situation" type

of question. When using it, however, one must remember that the extent to which one transfers learning always has some outer limit; and, consequently, the separation between the original learning situation and the test situation should be reasonable. (See 2, pp. 48-54, for a good discussion of this point.)

3. Essay questions should call for truly personal answers.

For most purposes, questions so framed that what the teacher wants is obvious, those which call out a desire to please the teacher, those which stir the "unfettered artistic imagination," or questions where there is external compulsion to answer in a certain way (either literal compulsion or compulsion through appeal to prestige, conventions, or morals) will furnish data of limited validity. If we are seriously seeking evidence bearing on fundamental changes in students, then questions which we ask them must be accepted as personally "real," "practical," and "honorable." Furthermore, the questions must give the impression of being susceptible of adequate treatment in terms of the student's own set of values. Testing here is very intimately tied up with teaching. The teacher who is too prone to express his own views in a forceful answer or one who dogmatically tells his students what is what on many issues, can expect little other than memory for what the teacher has said to be revealed through free responses to these issues.

4. The questions should be framed so as to encourage the testee to use his own "frame of reference," to reveal the "path of reasoning," to show reasons for choice, and defend positions taken.

This is another way of saying that the emphasis in essay testing should be on judgment rather than on memory, on *how* and *why* rather than on *what*, on *process* rather than *product*, on *means* rather than *ends*. What? Where? When? Who?: These are questions to be answered objectively. Explain, compare, contrast, evolve, assume, evaluate, justify, trace, prove, formulate: These words suggest situations which may involve true essays. (Although, as has already occurred to the reader, even these may call for nothing more than rote or semi-rote reproductions of materials to be given back to the teacher.)

5. Reliability should be sought through "depth" sampling and through paying particular attention to the representativeness of the samples.

The assumptions underlying the principle commonly accepted in objective testing that reliability is secured through using a relatively large number of independent items would not seem to hold for essay testing if it is used as a projective technique. Instead, it seems that the kinds of evidence on learning we seek through essay testing would be more dependably revealed through encouraging the student to "dig in" on a relatively few subjects rather than to react superficially to a large number of questions. This is not to imply, of course, that reliability in measurement is not important in essay testing. In fact, the variety of inferences which may be drawn from essay data (as compared with the single inference typically drawn from objective tests concerning the normative position on a total score) demands that particular attention should be paid to reliability. It does mean, however, that we cannot uncritically transfer to essay testing the "principles" determining reliability which have been evolved in objective testing. The emphasis probably should be placed on the representativeness of the sample. Experience in other fields with limited samples carefully chosen offers encouragement. In opinion polling, for instance, Roper in the *Fortune* polls, has demonstrated that relatively small samples which are highly representative may be more reliable than large samples not so well chosen.

III. *Suggestions Relating to the Handling of Answers to Essay Questions*

The reader who has followed our discussion to this point will realize that in the case of answers to essay questions one is dealing with highly complex data, and data which require considerable skill in handling if dependable inferences concerning learning are to be drawn from them. Furthermore, it is with some hesitation that the writer offers any suggestions in a field that has been subjected to so little experimentation. Perhaps the most appropriate recommendation would be to the effect that we should deliberately and systematically investigate the

problem of interpreting (that is reading, grading, or scoring) the essay examination as a projective technique. Experience with projective methods elsewhere, however, does indicate that the following suggestions are worth consideration:

1. To the extent possible, in evaluating essay answers we should use an inductive method; that is, we should examine the data and then evolve a "frame of reference" from which to treat them.

The reader's own purposes and values are sure to play a part in determining the plan finally developed for handling the data (as is true in evaluating any measurement), but their role will be a more legitimate one if they are introduced after a preliminary examination of the papers. One who approaches a set of essay answers with a fixed conviction as to what he *should* find, is defeating, in part at least, the legitimate purpose which the test serves. In fact, one might venture the suggestion that when one can legitimately predetermine the right answer one is dealing with learning that should be tested for by objective means.

2. In reading essay examinations, look for and carefully distinguish between the manifest and the "hidden" content; between data which reflect learning resulting from particular curricular experiences (influence of the course or courses) and those which reflect the general personality of the student; and between the data themselves (the answers) and the inferences which one draws from them.

The content, conceptualization, and insights consciously written into a paper are obviously important; but such matters as the style, omissions, rationalizations, or attitudes which are unintentionally introduced into the treatment may be just as revealing. If tests are designed to furnish evidence relating to the outcome of some particular learning experience (as is true in most of our teacher-testing) one must recognize, however, that not all of the evidence revealed through a personal document such as an essay answer is relevant to the task at hand. If the question is a real and vital one to the testee, then his total personality enters into his answer; and in achievement testing it commonly becomes the job of the reader to sort out that

which is the result of specific learning from that which is a reflection of his general personality only. Finally it would appear extremely easy for the careless reader to confuse the evidence (what the student actually writes) with the interpretations (what the reader infers from the answers). To do so can only encourage one in "going beyond the data" to an extent that is unjustified.

3. If other readers are to agree with one's evaluation of essay answers, in fact, if at a later date one is to agree with one's own interpretations, it is necessary to, (a) identify in a clear-cut manner exactly what is to be looked for, and (b) develop an organized and systematic method for observing the data and drawing inferences from them.

It is only through the above process that we get the agreement of judges (objectivity) in any psychological measurement. In the highly complex and revealing data which one gets in essays this is not the simple matter that it is in short-answer testing, where it is relatively easy to get acceptance, often uncritical acceptance, of our "rules for scoring and interpreting." Nevertheless this process is necessary if one is to justify his interpretations of essays to others. As the reader in all probability has already guessed, the writer does not subscribe to the current fetish for objectivity in teachers' measurements. A part of the disagreement among "competent judges" is entirely justified for the simple reason that different teachers commonly seek entirely different outcomes through their instruction. Another large part of the disagreement results simply from the fact that readers "play the game by different rules," a fact which may or may not be justified. However, it is only through repeating the examination of and interpretations of answers under the same conditions (that is, where readers look for the same things and handle the data in the same manner) that we are able to get information on the errors in our interpretations, and therefore, to get evidence on the dependability of our evaluations.

4. Normally, essay papers should be read for the information which they reveal concerning the learning of individual students.

The outcome of essay testing will usually be clues to and suggestions for further work with particular students. Perhaps it may be desirable for the reader to prepare a summary statement or a profile for each student, including the significant points in the answer and the inferences drawn from them. Certainly, we should avoid simple summations or a "counting up" of objective facts. To do so obviously hides the most important information revealed by the tests.

5. If grades must be determined from an examination of the papers, it is probably best, after reading, to identify the modal pattern of response and to classify the deviates into categories ranging along some continuum (or continua).

The writer's own experience with essay tests leads to the conviction that in any considerable group of essay answers such categories do exist and can be arranged in a hierarchy to which grades may be assigned. Sometimes it may be necessary or desirable to classify the answers more than once and along different continua, in which case grades may be arrived at through a process of averaging. When doing this averaging, it is usually desirable to give added weight to the extremes. (The writer has used, arbitrarily but to advantage, weights of 10, 7, 5, 3, 0 in a 5-fold classification.) In some cases it may be possible to develop a check sheet, a rating scale, or a set of objective questions for the use of the reader in evaluating each essay. Such devices are susceptible to quantification and have been found useful in getting a more dependable summary evaluation of complex behavior or the products of work in other fields. For ordinary purposes of grading, however, where only crude classification is demanded, the labor involved in developing such instruments would scarcely seem to be justified.

(These suggestions are not to deny, of course, the possibility of a supplementary treatment of essay data in some such objective manner as that reported by Stalnaker as being in use by the College Entrance Examination Board (5).)

6. And in conclusion, a word of caution: Beware of too broad generalizations.

Essay answers are records of behavior which we propose to interpret as signs of "characteristic ways of behaving." By

the nature of things, however, they are sure to be limited samples; and, therefore, conclusions drawn from them about the psychological make-up of students must be tentative in nature and subject to further study. Perhaps the most common criticism made of projective testing in general results from the tendency of many users of these methods to go too far beyond their data. When the methods are to be used to get information bearing on learning, there is an added complication resulting from the fact that in the early stages of one's learning behavior is apt to be particularly erratic and inconsistent. In consequence, in achievement testing, particularly, dependable generalizations can only be arrived at through an accumulation of signs pointing in a given direction.

REFERENCES

1. Allport, G. W. *The Use of Personal Documents in Psychological Science*. Social Science Research Monograph, No. 49, 1942
2. Findley, W. G. and Scates, D. E. "Obtaining Evidence of Understanding." *Measurement of Understanding*, 45th Yearbook. New York: National Society for the Study of Education, 1946. Part I, chapter IV.
3. Frank, L. K. "Projective Methods for the Study of Personality." *Journal of Psychology*, VIII (1939), 389-413.
4. Sargent, Helen. "Projective Methods: Their Origins, Theory and Application in Personality Research" *Psychological Bulletin*, XLII (1945), 257-293.
5. Stalnaker, J. M. "Essay Examinations Reliability Read." *School and Society*, XLVI (1937), 671-672

THE NATURE OF VERBAL FLUENCY¹

BENJAMIN FRICHER
University of Southern California

I. Introduction

DURING World War II interest was aroused in the effect of fluency on reaching correct solutions to judgment problems which arise in connection with military flying. Questions to which answers are needed include: (a) Is fluency an element of the judgment process? (b) Is fluency confined to verbal situations or is it a broader function operating in both verbal and non-verbal media?

A possible answer to the latter question seemed to be contained in Thurstone's *Primary Mental Abilities Battery* (8). Thurstone analyzed a battery of fifty-seven tests into thirteen factorial components. Of the nine factors that he was able to identify, one was interpreted as a word-fluency factor. Analyses performed since have revealed other fluency factors in tests similar to those analyzed by Thurstone (3, 6, 7). It was felt that, since not all of Thurstone's axes had been rotated to positions of simple structure and positive manifold (9), and since some of his factors were not identified, a sub-battery of tests from the *Primary Mental Abilities Battery* might reveal additional fluency factors. Also, further information concerning the factorial composition of tests included in Thurstone's battery has been obtained in several subsequent analyses, performed during World War II, which indicate that (a) vocabulary tests should have higher loadings on the verbal factor than they have in Thurstone's solution (1), and (b) the spatial-relations and visualization variances should be brought out on separate factors (1). All of these considerations indicated that

¹ This paper is a revision of a thesis presented by the author as a candidate for the M.A. degree at the University of Southern California, 1946.

an analysis of a sub-battery of Thurstone's tests might further clarify the problem of the nature of fluency if the final rotated positions of the axes were determined by the principles of simple structure and positive manifold.

In this study fluency is defined as the ease or speed of evocation of original associations and hypotheses, both verbal and non-verbal. Woodworth, in his summary of the experimental literature on the higher mental processes suggests that there may be two fundamental processes involved in solving all problems: (a) calling up various acts or responses and (b) selecting those responses which meet the requirements of the problem (11). Fluency would be identified with the first of these categories. Some English psychologists, among them Brown and Stephenson, use the term fluency to describe a personality trait which they measure by the degree of uninhibitedness and extraversion (2).

II. *Literature on Verbal Fluency Factors*

Thurstone's Word-Fluency Factor.—In the first major application of Thurstone's method of factor analysis, 57 psychological tests were given to 240 male college students. One of the meaningful factors was identified as a word-fluency factor and has the following tests:

<i>Test</i>	<i>Loading</i>
Anagrams534
Grammar530
Disarranged Words512
Spelling508
Vocabulary (Thorndike)413
First and Last Letter388

Thurstone identifies the principal characteristic common to these tests as a fluency in dealing with single and isolated words and contrasts it to dealing with ideas and meanings (8).

Thurstone and Thurstone's High-School Study.—Thurstone and Thurstone made a second study with 268 high-school seniors as subjects. Thirty-six tests were used. Again a word-fluency factor was identified and described as the ability to think of isolated words at a rapid rate. They discuss their results as follows:

One of the most interesting findings in the present study is the distinction between the verbal-comprehension factor (V) and the word-fluency factor (W) which can be illustrated by two of the tests in the present battery. Both of these tests involved synonyms, but the test procedures were different. One of these tests, Vocabulary, was an ordinary recognition form of vocabulary test in which the subject merely checked the response word which had the same meaning as the given stimulus word. This test had high saturation (.68) on the V factor, which accounts for half the variance of the test. Its saturation on the W factor vanishes, (.015). This is as we should expect, because the response words are printed in the test so that the subject need only check the correct word. The test involves the understanding of the given words, but the subject is not required to supply words. In order to feature the W factor, we included a free-recall form of synonyms test. The test gives a list of common adjectives which are easily understood by all of the students. They are asked to write three synonyms for each given word. Here it is not a question of whether the student subjects understood the given words, since they were all ordinary adjectives. The task was to supply quickly three synonyms for each given word. The test had a high saturation on the W factor (.51) but its saturation on V vanished (.10).

Thornton's Persistence Battery.—Thornton analyzed a battery of persistence tests to detect a general persistence factor. The subjects were 189 college students in a beginning psychology course. No general factor appeared. Five orthogonal factors were obtained and identified as (a) withstanding discomfort, (b) keeping on at a task, (c) sex difference (or strength), (d) feeling of adequacy, and (e) verbal fluency. He believed the last factor to be between Thurstone's two verbal factors (V and W). The tests Thornton used to identify the verbal-fluency factor are:

<i>Test</i>	<i>Loading</i>
Verbal Ability712
Reading Speed634
Word building speed589
Number of words built382
Amount read on perceptual test	.381

Thornton makes the following suggestions for the further investigation of verbal fluency:

A question was raised earlier as to whether the fluency factor is verbal-fluency only. This question could be answered experimentally by a factor analysis of a battery of tests chosen so as to include some verbal and some non-verbal tasks which would give opportunity for the fluency factor to influence the score.

A more basic question concerning the fluency factor is whether it is an ability or a predisposition or a combination of the two. Some subjects appear to have a predisposition to work rapidly and intensely at a task. Is this apparent predisposition simply a reflection of the ability to work rapidly, or is it a separate factor? (7)

Carroll's Verbal Abilities Battery.—Carroll analyzed a battery of 42 tests in the domain of speech and language behavior. The subjects were 119 college students. He found Thurstone's *V* factor split into two, possibly three factors. He identified two of these as "Richness of the individual's stock of linguistic responses" and "The ability to handle semantic relationships." No satisfactory interpretation was found for the third factor.

Thurstone's *W* factor seemed to split into two component factors. Carroll advanced the hypothesis that one of these factors involves speed-of-word association (usually for common words) where there is some element of restriction in the task imposed, i.e., where only one or a certain number of responses from the total reserve are correct. In *Disarranged Words*, for example, the test materials undoubtedly give rise to a number of implicit responses from which the subject must select the correct or acceptable responses. Other tests having significant projections on this factor are *Suffixes*, *Form-naming*, *Word-number Memory* and *Color-naming*.

The other factor seems to involve, in some way, the rate of production of meaningful and syntactically coherent discourse where there is little restriction to definite responses. *Theme-word Count*, the test with the highest projection on this factor, clearly involves facility in producing sentences which are sufficiently meaningful to be accepted by the subject. The other tests having appreciable projections on this factor, *Grammar*, *Similes*, and *Picture Description* (per cent of relevant words) seem also to involve an element of syntactical coherence. Tests having significant loadings on this factor are *Suffixes*, *Form-naming*, *Word-number Memory*, and *Color-naming* (3).

Johnson and Reynold's Verbal Abilities Battery.—Johnson and Reynolds analyzed problem-solving in verbal tasks. They started with the assumption that there are two fundamental types of activity in problem-solving: (*F*) the flow of various acts or responses, and (*S*) the selection of these responses according to the requirements of the problem. They selected a battery of ten verbal tests which they believed included *F* and *S* in varying amounts and administered them to 113 college students. The test scores were then correlated and factored by Thurstone's centroid method. Two factors were obtained. These were identified as *F* and *S*. Table 1 gives the rotated factor loadings for this battery.

TABLE 1
Final Factor Pattern of Johnson and Reynold's Verbal Ability Battery

Test	<i>F</i>	<i>S</i>	<i>h</i> ²
1. Flow564	.024	.319
2. Multiple Completion488	.135	.256
3. Paired Opposites340	.246	.174
4. Word Cross-outs418	.564	.492
5. Same and Opposites204	.818	.710
6. Verbal Relations226	.795	.682
7. Nelson-Denny Vocabulary143	.776	.622
8. Restricted Completion067	.677	.462
9. Henmon-Nelson Intelligence230	.763	.634
10. Nelson-Denny Reading516	.575	.601

Test 9, the *Henmon-Nelson Intelligence Test*, indicates the relative saturation of these two factors in an intelligence test (6).

Cattell's Surgency-Desurgency Factor.—Cattell identified a Surgency-Desurgency factor as one of his twelve primary personality factors. He discusses the relationship of this factor to verbal fluency as follows:

A test factor—fluency of association—has been in use for some time as a measure (Validity = 0.65) of the cluster surgency. Now that the cluster has been resolved into two and possibly three factors, it remains to be seen if the present factor, as one might suspect, will carry the main correlation with fluency. Fluency loads: Verbal ability (very easy opposites) .71, Speed of reading .5–.6, Associations in ink blots .5–.6.

The writer has unpublished experiments on three closely studied cases of cyclical depression in which great changes in fluency score accompanies mood change. Everything points to the manic-depressive—or, at least the depressive-normal, change being directly along this F axis (4).

III. *Factor Analysis*

A sub-matrix of twenty tests was selected from Thurstone's battery of fifty-seven tests. Tests were selected for one or both of the following reasons: (a) they seemed to require the rapid evocation of original hypotheses and ideas and therefore were presumed to have a "fluency" component and (b) their factorial composition is well known and was used to locate and identify common factors such as spatial, reasoning, verbal, perceptual. This was done so that if any new fluency factors should appear they would be clearly distinguishable from these older, better established factors.

The inter-test correlations used for the analysis are those computed by Thurstone.² They are tetrachoric correlation coefficients based upon a sample of 218 college students, divided as follows: 132 Freshmen, 36 Sophomores, 30 Juniors, 12 Seniors, and 14 college graduates. The modal age was 18. Scores were available for 113 of the subjects on the *Psychological Examination of the American Council on Education*. The mean score of the group was found to be 1.63 standard deviations above the national norms for 203 American colleges, indicating a sample with superior mental endowment.

The correlations are positive or zero with but two exceptions: between Controlled Associations and Cubes ($r = -.10$) and between Controlled Associations and Flags ($r = -.16$).

It was thought advisable to compute centroid factor loadings based on the sub-matrix rather than to use the centroid loadings given by Thurstone. His loadings were computed from a matrix of fifty-seven tests and more factors may have been extracted by him than could be accounted for by the present battery of twenty tests.

² For descriptions and sample items of the tests see Thurstone, L.L., "Primary Mental Abilities," *Psychometric Monograph*, I (1938), chapter II. For inter-test correlations, see Appendix A.

TABLE 2
Centroid Loadings, Communalities, and Reliabilities of the Tests

Test	I	II	III	IV	V	VI	VII	VIII	μ^2	r^2
1 Figure Classification53	.39	.19	-.26	.16	-.04	-.13	-.06	.61	.90
2 Controlled Association39	-.48	-.11	-.20	.26	-.31	.22	.06	.65	.90
3 Inverse Opposites69	-.48	-.10	.06	-.15	-.13	.10	-.07	.75	.91
4 Completion80	-.34	-.21	.27	-.17	-.13	-.11	.15	.95	.84
5 Disarranged Words63	-.22	.18	-.07	.28	.19	-.19	.13	.65	.95
6 First and Last Letter59	-.27	.36	-.16	-.07	-.10	.18	.10	.63	.68
7 Anagrams48	-.25	.40	-.19	-.16	-.08	.16	-.04	.55	.86
8 Cubes64	.49	.27	.29	-.09	-.12	-.22	-.04	.88	.96
9 Flags51	.49	.23	.26	.13	.19	.12	-.07	.69	.97
10 Form Boards80	.35	-.13	.13	-.09	-.09	.30	.19	.94	.97
11 Punched Holes66	.39	-.15	-.17	.09	.22	.18	.17	.76	.94
12 Mechanical Movements65	.20	-.20	-.26	-.26	.09	.03	.15	.67	.91
13 Identical Forms45	.05	-.10	.20	.10	-.12	-.33	.20	.43	.98
14 Pursuit48	.43	.23	.17	.18	-.23	.22	-.18	.66	.98
15 Arithmetic Reasoning62	.13	-.22	-.10	-.16	.16	.16	-.25	.60	.69
16 Verbal Analogies77	-.03	-.11	-.04	.12	-.12	-.05	-.19	.67	.97
17 Pattern Analogies76	.14	-.12	-.17	-.16	-.22	-.20	-.18	.79	.58
18 Vocabulary60	-.53	-.32	.34	.21	.26	.17	-.12	1.01	.96
19 Free Writing23	-.21	-.05	-.11	.32	.08	-.23	.12	.29	.29
20 Vocabulary (Thorndike)79	-.29	-.14	.10	-.23	.27	-.13	-.20	.92	.84

Nine factors were extracted but only eight were used in the rotations because the loadings on the ninth factor appeared, upon inspection, to be insignificant, ranging only from $+.15$ to $-.12$. Fewer factors may be expected to be found in this solution than in Thurstone's, not only because fewer tests are involved, but also because the variables were selected with some pre-knowledge of their factorial composition and purity.

The extraction of centroid factors was made according to the procedure outlined by Guilford (5).

Table 2 presents the centroid loadings, communalities and reliabilities of the tests. The reliabilities are corrected odd-even tetrachoric correlation coefficients.

The centroid loadings shown in Table 2 were plotted on graph paper two axes at a time and rotated with the principles of simple structure and positive manifold in mind. These principles call for maximizing of the number of zero or near-zero loadings and minimizing of the number of significant negative loadings.

Also followed was the practice of placing axes for well-established factors in the positions where numerous previous analyses have shown them to belong. This was done on the supposition that if all of the factors but one in an analysis are of a well-established nature, they will determine the location of the new axis, and to that extent the location of the new axis will be more objectively determined.

A new graphical method of rotation which reduces time and labor was employed (12). Twenty-four rotations of pairs of axes were necessary to satisfy the criteria of simple structure and positive manifold. Loadings of less than $.2$ were considered as vanishing entries. Table 3 gives the final rotated factor loadings and the communality of each test. In no case did the communality of a test after rotation differ from its communality before rotation by more than $.02$.

IV. *Interpretation of the Factors*

Inspection of the rotated factorial matrix (Table 3) reveals that, in the main, positive manifold and simple structure have been attained. We may now consider the interpretation of the

TABLE 3
Final Rotated Factor Loading

Test	A	B	C	D	E	F	G	H	<i>h</i> ²
1. Figure Classification10	.49	.16	.37	.26	.20	-.13	.26	.60
2. Controlled Association16	.14	.13	.15	.00	.00	.53	.53	.65
3. Inventive Opposites53	.00	.01	.26	.10	.22	.54	.18	.73
4. Completion62	.11	.07	.16	.47	.17	.52	.07	.95
5. Disarranged Words42	.03	.05	.23	.31	.55	.06	.03	.64
6. First and Last Letter19	.11	.18	.14	-.02	.56	.35	.31	.63
7. Anagrams14	.11	.07	.20	-.11	.57	.29	.21	.54
8. Cubes09	.77	.07	.14	.39	.27	.06	-.14	.87
9. Flags20	.72	.29	.00	.04	.13	-.15	.01	.68
10. Form Boards30	.50	.62	.24	.24	.09	.30	-.06	.94
11. Punched Holes25	.34	.63	.33	.20	.05	-.12	.14	.76
12. Mechanical Movements . .	.25	.11	.45	.50	.30	.20	.04	-.05	.66
13. Identical Forms19	.25	.02	.01	.55	-.01	.13	.12	.43
14. Pursuit	-.03	.74	.22	.11	-.05	.05	.20	.08	.66
15. Arithmetic Reasoning . .	.42	.26	.25	.53	-.02	.04	.05	-.03	.59
16. Verbal Analogies36	.37	.08	.43	.22	.05	.27	.30	.67
17. Pattern Analogies19	.36	.06	.64	.36	.15	.25	.07	.79
18. Vocabulary89	.04	.03	.00	.00	-.14	.28	.32	1.00
19. Free Writing15	-.06	-.03	.01	.26	-.01	-.04	.44	.29
20. Word Knowledge77	.14	-.05	.40	.21	.26	.13	.06	.91

rotated factors. Since the correlation coefficients on which this analysis is based are relatively unstable, only projections at least as large as $\pm .4$ will be considered significant.

The variables having loadings of .40 or greater on factor A are ranked below in order of size of projection. Significant projections of these tests on other factors are also given.

Test	Projections	
	A	Other Factors
18. Vocabulary (58) ^a89
20. Word Knowledge (60)77	.40 D
4. Completion (11)62	.47 E; .52 G
3. Inventive Opposites (10)53	.54 G
5. Disarranged Words (12)42	.55 F

The large number of tests having significant loadings and the large amount of the loadings of some of the tests make identification of this factor relatively certain. The *Vocabulary Test* which has approximately 79 per cent of its variance accounted for on this factor, is a fairly difficult vocabulary test

^a The number in parentheses following the test name is the variable number in Thurstone's *Primary Mental Abilities Battery*.

of the multiple-choice type. The *Word-Knowledge Test* is a similar test at a somewhat lower level of difficulty. It has approximately 60 per cent of its variance accounted for by this factor. The other tests having significant loadings on this factor also seem to have a knowledge of the *meaning of words* as an important part of the task, whereas none of the tests with vanishing loadings seems to require this knowledge. The factor can therefore be identified with reasonable confidence as the verbal factor which has been identified previously in a number of studies.

Factor B has the following tests:

<i>Tests</i>	<i>B</i>	<i>Other Factors</i>
8. Cubes (18)77	.39 E
14. Pursuit (27)74
9. Flags (20)72
10. Form Boards (21)50	.62 C
1. Figure Classification (8)49

None of the tests with loadings of .4 or higher have verbal content. The tests with the highest loadings seem to involve the ability to see spatial relationships in two or three dimensions. This function seems to be absent from all tests with negligible loadings. This factor can be identified with the spatial-relations factor found in numerous previous analyses.

The tests with significant loadings on factor C are:

	<i>C</i>	<i>Other Factors</i>
11. Punched Holes (24)63
10. Form Boards (21)62	.50 B
12. Mechanical Movements (25)45	.50 D

The common function here seems to be the ability to visualize. Evidently Thurstone had both the spatial and the visualization variances of his tests appear together on the same factor. A clue to the difference between these functions may be obtained by examining the tasks involved in Cubes, which has high spatial saturation and no visualization saturation, and in Form Boards, which has high loadings on both factors. In

the *Cubes Test* the subject is asked to recognize whether the parts on the sides of the cube patterns hold the same relationship to each other after the cube has been rotated. Both original and final positions are shown and need not be visualized. In the *Form Boards Test*, however, not only must the parts given be visualized in a new, rotated relationship, but the final positions of the parts within the form must likewise be visualized.

The tests high on factor D are:

<i>Tests</i>	<i>D</i>	<i>Other Factors</i>
17. Pattern Analogies (44)64
15. Arithmetic Reasoning (39)53	.42 A
12. Mechanical Movements (25)50	.45 C
16. Verbal Analogies (41)43
20. Word Knowledge (60)40	.77 A

The common function that the tests high on this factor seem to share is reasoning ability. It is not clear, however, why one vocabulary test should have a significant loading on this factor and the other none at all.

Significant loadings on factor E are:

<i>Tests</i>	<i>E</i>	<i>Other Factors</i>
13. Identical Forms (26)55	.. .
4. Completion (11)47	.62 A; 52 G
8. Cubes (18)39	.77 B

It is customary to identify the factor on which the loading of identical forms or similar tests appears relatively pure as a speed-of-perception factor. It is interpreted as the ability to perceive quickly details imbedded in irrelevant material. It is not clear, however, why a test which is wholly verbal in content, like Completion, should have a significant projection, whereas other tests, more obviously visual, do not. It may be that, due to the low difficulty of the items of the *Completion Test*, speed of reading was a significant factor in determining final scores.

Factor F has the following tests:

<i>Tests</i>	<i>F</i>	<i>Other Factors</i>
7. Anagrams57	...
6. First and Last Letter56
5. Disarranged Words55	42 A

Thurstone's description of the function underlying this factor as a fluency in dealing with words, separate from their ideas and meanings, still seems satisfactory (8). The absence from this factor of other verbal and non-verbal tests which presumably have an element of fluency indicates word fluency to be a restricted type of fluency.

Factor G has the following tests:

<i>Tests</i>	<i>G</i>	<i>Other Factors</i>
3. Inventive Opposites (10)54	.53 A
2. Controlled Associations (9)53	.53 H
4. Completion (11)52	.62 A; 47 E

This factor is restricted to the three tests which were selected to represent "speed of calling up pertinent associations." It can probably be identified with Carroll's speed-of-word-association factor and Thornton's mental-fluency factor (3, 7). Here, as in factor F, significant loadings for non-verbal tests are conspicuous by their absence and the function represented seems to be restricted to verbal tasks.

Only two tests have significant loadings on factor H.

<i>Tests</i>	<i>H</i>	<i>Other Factors</i>
2. Controlled Associations (9)53	.53 G
19. Free Writing (59)44

Since this factor is what is commonly referred to as a "doublet," i.e., only two tests on it have significant loadings, it is not possible to tell whether some function specific to these two tests or a broader function is represented. In the latter case some interesting possibilities present themselves. One is that it is a "speed of writing" factor. Another possibility is that the com-

mon function is "verbosity," the speed of producing a large number of words in a limited period of time. In either case, it would seem to be related to Johnson and Reynold's "free-flow-of-responses" factor and Carroll's "speed-of-production-of-meaningful-discourse" factor (3, 6).

V. Discussion

In addition to a word-fluency factor, Thurstone's data yielded two other factors which resemble factors previously identified as fluency factors. One of these is interpreted as a speed-of-association factor. It resembles Thornton's verbal-fluency factor, Carroll's speed-of-association factor, and Johnson and Reynold's flow factor (3, 6, 7). The crucial function involved seems to be the ability to associate rapidly a large number of words bearing a given relationship to a stimulus word.

The other candidate for interpretation as a fluency factor is a doublet and it cannot, therefore, be identified in the present solution. It does, however, bear a resemblance to Carroll's "Rate-of-production-of-meaningful-and-syntactically-coherent-discourse" factor (3). The number of words written in a theme seems to be a relatively pure measure of it. If this factor can be verified, it probably comes close to what is usually meant by fluency of organized verbal discourse.

There is no indication in the results that fluency is a generalized function appearing in several media. There is, in fact, very little overlapping of any function in verbal and non-verbal tests. The reasoning factor is the only exception, three verbal and two non-verbal tests having significant projections on it.

The following are the conclusions from this study:

1. There are two, possibly three, independent types of verbal fluency.

- a) Fluency in dealing with single words.
- b) Fluency of association for common words where there is some restriction placed upon the response.
- c) Rate of production of syntactically coherent discourse.

2. Vocabulary tests in Thurstone's *Primary Mental Abilities Battery* are more heavily saturated with verbal content than he had found.

3. It is possible to bring out separate spatial-relations and visualization factors from Thurstone's data, replacing his one spatial factor.

4. Verbal and non-verbal tests tend to be projected on separate factors, the reasoning factor appearing to be an exception.

5. No evidence of a universal fluency factor, common to verbal and non-verbal tests, was found.

Summary

In view of the following considerations, it was believed that an analysis of a sub-battery of Thurstone's *Primary Mental Abilities Tests* might yield further information concerning the nature of fluency:

- (1) One of the factors identified by Thurstone was a word-fluency factor.
- (2) Thurstone's solution included some unidentified factors.
- (3) *The Primary Mental Abilities Battery* contains tests similar to some that were found to define fluency factors in other analyses.
- (4) Evidence from subsequent analyses indicated that vocabulary tests should have higher verbal loadings than they do in Thurstone's solution and that separate visualization and spatial factors should replace Thurstone's spatial factor.
- (5) Not all of Thurstone's axes had been rotated to positions of simple structure and positive manifold, indicating that another solution might be found which could shed more light on the nature of fluency as well as resolve the problems mentioned above.

A sub-battery of twenty tests was selected from Thurstone's *Primary Mental Abilities Battery*. Eight centroid factors were extracted and rotated, two at a time, by a new rotational technique, to psychologically meaningful positions.

Six of the factors were identified with reasonable confidence. They are verbal, spatial-relations, visualization, reasoning, perceptual, and word-fluency. Two factors are less well identified. One has been interpreted as a "speed-of-association" factor. The last factor is a doublet and cannot be interpreted but seems

to bear a resemblance to a previously identified verbal-fluency factor.

The major conclusions from this study are:

1. Two, possibly three, types of verbal fluency can be identified.

2. A universal fluency factor, common to both verbal and non-verbal tests, did not appear.

3. Separate spatial and visualization factors can be brought out from Thurstone's data.

4. The vocabulary tests are more heavily saturated with verbal content than Thurstone had found.

REFERENCES

1. *Aviation Psychology Series, Report No. 5*. Washington: U. S. Government Printing Office, 1947.
2. Brown, W. and Stephenson, W. "A Test of the Theory of Two Factors." *British Journal of Psychology*, XXIII (1933), 352-370.
3. Carroll, John B. "A Factor Analysis of Verbal Abilities." *Psychometrika*, VI (1941), 279-308.
4. Cattell, R. B. "Interpretation of the Twelve Primary Personality Factors." *Character and Personality*, XIII (1944), 55-91.
5. Guilford, J. P. *Psychometric Methods*. New York: McGraw-Hill Book Company, 1936.
6. Johnson, Donald M. and Reynolds, Floyd. "A Factor Analysis of Verbal Abilities." *The Psychological Record*, IV (1941), 183-195.
7. Thornton, G. R. "A Factor Analysis of Tests Designed to Measure Persistence." *Psychological Monographs*, LI (1939), No. 229.
8. Thurstone, L. L. "Primary Mental Abilities." *Psychometric Monograph*, I (1938), 82-85.
9. Thurstone, L. L. *The Vectors of the Mind*. Chicago: University of Chicago Press, 1935.
10. Thurstone, L. L. and Thurstone, T. G. *Factorial Studies of Intelligence*. Chicago: University of Chicago Press, 1941.
11. Woodworth, R. S. *Experimental Psychology*. New York: Henry Holt and Company, 1938.
12. Zimmerman, W. S. "A Simple Graphical Method for Orthogonal Rotation of Axis." *Psychometrika*, XI (1946), 51-55.

PRACTICAL SUGGESTIONS FOR IMPROVING TESTING PROCEDURES

LEO CHALFEN
New York University

I. *Introduction*

THE concern of this paper is the development of an approach to a testing procedure which is applicable in most situations. Unfortunately, much desirable information is not readily accessible in the present state of orientation of testers and the present stage of development of psychological instruments. For the psycho-clinician this lack presents a threefold challenge to action: an advancement of research to aid in the construction of better diagnostic and prognostic instruments; a better integration of theory and practice; and a more thorough utilization of the knowledge and techniques which are available. The primary concern of this paper is with this last purpose.

A more fruitful use of present instruments is hampered by an ignorance of available procedures, by a lack of creative imagination, and by certain artificial limitations which are self-imposed. The concept of what is implied in the scientific method must be clarified. A more extensive use of the clinical criteria of representativeness, optimity, and appropriateness (see section II) is suggested both in obtaining and in assessing test results. The construction and use of test batteries and test sequences must be justified psychologically as well as statistically.

II. *Implications of the Scientific Method*

Unfortunately, many testers formulate hypotheses as to the status of the testee in a routine fashion on the assumption that there is a one-to-one correspondence between the behavior of the subject and the ability which theoretically underlies the be-

havior. This approach lacks scientific rigor. The tester should avoid making an assumption on this point. His hypothesis should be subjected to inquiry and experimentation without restriction, keeping in mind the fact that the test is an analysis of behavior at a certain moment and under certain conditions. At times the results are unexpected, but always accepted and systematized by the competent examiner. It should be stressed that hypotheses precede statistical manipulation and standardization. Spearman (7), after Udney Yule, says that the main function of statistics is to aid in testing conclusions already reached by other means. This point is vital to the appreciation of a method of testing which does not adhere rigidly to a fixed procedure in test administration.

Beck (1), Wells (9), and Hoist *et al.* (2) refute the implication that a test item has the same meaning for every subject. Differences in interpretation of items may be misconstrued as actual differences in the subjects. It is, therefore, the task of the examiner to present to the subject "... the test problems in those proportions which are closest to those by which the norm was established" (9). For example, the use of the *Worcester-Wells Memory Test* in a state other than Massachusetts would necessitate the substitution of local cities for those given in the test, and even then there would be no assurance of comparability.

Variables can be under some measure of control in testing so that one knows readily the relationship of stimulus and response. Assuming that the test is valid, that an intelligence test actually measures intelligence, or that a test of manual dexterity measures speed and accuracy of movement, the examiner is not free to introduce extraneous factors into the testing situation (and interpretation), such as persistence, anxiety, reading proficiency, or the ability to understand unintelligible instructions. The expert examiner controls a testing situation, at times rephrasing instructions or shifting the order of tasks within a test (as allowed on the Merrill-Palmer and the Wechsler-Bellevue scales (9, pp. 171-172)), so that the subject is aware of what is desired and is cooperating actively. Without such cooperation the subject may intrude intentional or unintentional

tional deception on self-administered personality and interest inventories as found by Steinmetz when using the *Strong Vocational Interest Blank*.

The observations of many testers have resulted in their stressing the deleterious effect on test performance of attitudes such as recalcitrancy, lack of self-confidence and nervous excitement. Several studies (3, 4, 6) have shown that psychometric results in the intellectual and psychomotor areas are affected by the attitude of the subject.

Problems arise which cannot be handled unless the best use is made of the available armamentarium of the tester. In illustration, note the problems of measuring the abstract ability of a subject with organic brain damage, the educational interest pattern of an adolescent girl, or the manual aptitude of a blind person. The first case might be handled by administering the Rorschach or the Hanfmann-Kasanin version of the *Vigotsky Test*, even though neither of these has as yet been adequately validated. The adolescent girl might be given the *Garretson-Symonds Interest Questionnaire* even though this was standardized only on high-school boys. The blind person's manual dexterity could be measured by some of the existing tests which do not depend primarily on vision, after enough practice trials are provided to give spatial orientation. The material so obtained can be of real aid if it is interpreted with full awareness of the limitations of the method.

III. Suggestions from Clinical Criteria

The meaning and usefulness of the results can be judged against the criteria of representativeness, optimity (4) and appropriateness. Representative results are elicited when the subject is cooperative (exhibits self-confidence, interest and effort). If these essentials are not present in an adequate amount there can be no assurance that the test results are indicative of the subject's true nature and of the level of behavior—as far as the tests can measure. Roc and Shakow (4) list other conditions outside the voluntary control of the subject and usually of a temporary duration which affect the representativeness of test results, such as physical conditions, fatigue, and physical

LIBRARY & DOCUMENTATION

UNIVERSITY OF CALIFORNIA

illness (headache), physical disability (loss of hearing aid), emotional upheaval, or some types of psychotic episodes (manic hallucinations). Hoist *et al.* (2) say ". . . performance in an activity can not be viewed as an isolated phenomenon outside the environmental context in which the activity takes place." It seems clear, therefore, that expressing results in any such terms as IQ or wpm. (in a typing test) is meaningless when they are not representative for either of the aforementioned internal or external reasons. However, in keeping with the thesis, they can be presented descriptively even though they are unrepresentative, granted that they are used as diagnostic elements.

Roe and Shakow (4) also mention the concept "optimity." Results are optimal only if they are representative, if the subject has no language or permanent physical handicap (blindness), is under forty years, and if the examiner considers the results indicative of the highest level the subject has ever attained. It seems sounder psychologically to limit this construct of optimity to measures of capacity (and to broaden the base of such a judgment to include the highest level of attainment predictable for the future). Optimity is judged partially by psychometric signs like consistency between scores of tests of the same dimensions, consistency of intrasubtest and intersubtest scores, and comparison of the results of repeated tests of the same function. It is judged partially against biographic criteria like occupational level and the nature of responsibilities ever held, the age of attaining the level of highest education, and the social and civic status throughout the years.

The third clinical criterion is that of the appropriateness of a test. To satisfy this criterion the examiner justifies the purpose for testing. The variables to be considered in choosing the right instruments for the task are dependent upon the nature of the problem and the availability of sound tests. For example, in measuring the intellectual level of a normal adult the Wechsler-Bellevue is generally preferable to the Stanford-Binet even though the latter is a valid measure of intelligence, namely, because the standardizing population of the former includes adults. Yet for an uncooperative adult mental patient the *Goodenough Draw-A-Man Test*, developed for use with

children, may be profitably administered. In various situations different variables have relative importance in determining the choice of test; these are the testee's age, sex, educational and cultural background, and the type of information desired.

IV. Selection and Use of Test Batteries and Sequences

A procedure about which little is known or done is the construction of test batteries and sequences. A test battery is a series of tests which is dependent upon many factors for its content, extensity, intensity, direction and sequence. A rationale covering these points may be established upon principles of testing, efficiency and feasibility. The actual content of the test material is determined by the original testing problem and its development as the session progresses. The compilation and collation of the various measures used for individual test batteries is dependent first on their reliability, validity and appropriateness.

The number of tests used gives little indication of the extensity of measurement. Seidenfeld (5) deplors the "shotgun test procedure" by which the subject is exposed to innumerable tests in the hope that some worthwhile information will be gained. Professional competence and sound clinical judgment is invaluable in the choice of a functionally optimal number of measures in the battery adequate for the task. The psychological effect of being bombarded with an excessive number of tests is not definitely known, but it is quite unlikely to be conducive to feelings of confidence and interest and effort. Intensity of measurement can sometimes be subjected to the criticism of unnecessarily duplicating measures plus that of overstepping the bounds set by the testing problem, the policy of the institution or the capability of the examiner.

The matters of sequence and direction are relatively unknown factors. What is the effect on the subject and on his results of any one test following another? Since experimental data does not define this area, it is suggested that greater reliance be placed on the clinical judgment of the examiner. Results which are more likely to be representative and optimal as well as reliable and valid may be obtained by satisfying the

criteria of section III. This means planning a sequence so that the subject does not become bored or fatigued or unduly anxious during tests not primarily intended to measure anxiety, or uncooperative because of apprehension or lack of rapport. To avoid these pitfalls—as far as the tests are concerned—the sequence could well be planned on an individual basis in an attempt to comply with the purpose for the testing and to adapt the means to the subject. Quite probably the best procedure will be one which begins with a measure which is meaningful (in relation to his problem) to the subject. The use of some brief obvious test of ability like a spelling test for clerical workers or a rate of manual manipulation test for packers and shippers often solves this problem. In the early part of the battery the subject is less likely to react unfavorably to those tests which may tax his patience or endurance. If many measures requiring concentrated effort are to be given, interspersal between tests which are untimed or which differ in the function measured may attenuate the fatigue. For example, if two tests of mental ability are used, they should not follow one another any more than two measures of finger dexterity. In using personality tests it seems better to precede the fairly obvious inventories with the neutral or ambiguous content projective techniques. If procedures which have charged content or which present stress situations are left for the latter part of the testing session or for another period, there would be less information lost in case of any reason for discontinuing the testing.

The administration of a series of tests is further complicated by planning the direction of the series to measure those aspects of the psychological dimensions (intelligence, interest, personal and social adjustment, capacity, proficiency) which are pertinent to the testing problem as it develops.¹ In illustration, to help plan a practical vocational goal, the early definition of pertinent dimensions (intelligence and vocational interest) is desirable. From the results obtained probing need be done only in those areas where planning is feasible (this approach

¹ A good illustration of directional planning is seen in the program of the Test Division of the Psychological Corporation devised for the Community Advisory Service Center of Bridgeport, Connecticut.

meets the obligations of the criterion of efficiency of testing inherent in Seidenfeld's article (5)). An individual of borderline intelligence cannot plan on a professional career, nor will a subject whose measured interests are in literary and persuasive pursuits be most satisfied in manual activity. In a diagnostic problem the same procedure holds, early definition of pertinent dimensions to aid in the construction of an efficient, feasible test battery (such as, intelligence and visual ability in a reading difficulty; personality structure and dynamic content in a differential diagnosis between a reactive depression and a manic-depressive depression).

Frequently the course of the testing changes as new information appears. For example, the original problem may be the advisability of the subject's considering engineering as a career. This plan may be ruled out immediately if his intellectual level is too low, at which point the testing may be oriented toward related occupations requiring a lower level of intelligence. If his measured interests were in social welfare and his evaluative attitudes high in the social and religious spheres the direction of the testing could be changed accordingly. If an institutionalized patient were being tested, and his intelligence test results suggested brain damage, specific study of this problem would certainly follow. Direction and sequence cannot be planned separately according to the preceding discussion.

V. Summary

In an attempt to obtain meaningful and useful information from tests it is suggested that current tools be utilized to their fullest extent by psycho-clinicians. Test information may be more useful when elicited and evaluated against the clinical criteria of representativeness, optimity, and appropriateness. Test battery and test sequence administration may be profitably based on clinical judgment, at least until experimental data define this area.

REFERENCES

1. Beck, S. *Rorschach's Test: I, Basic Processes*. New York: Grune and Stratton, 1944.

2. Horst, P. *et al.* *The Prediction of Personal Adjustment*. New York: Social Science Research Council, 1941.
3. Huston, P. E., Shakow, D. and Riggs, J. A. "Studies of Motor Function in Schizophrenia; II, Reaction Time." *Journal of General Psychology*, XVI (1937), 39-82.
4. Roe, Anne and Shakow, D. "Intelligence in Mental Disorders." *Annals of the New York Academy of Science*, XLII (1942), 361-490.
5. Seidenfeld, M. A. "On 'Testophilia' as a Disease of Professional Psychologists: An Editorial." *Journal of Clinical Psychology*, I (1945), 343-344.
6. Shakow, D. and Huston, P. E. "Studies of Motor Function in Schizophrenia; I, Speed of Tapping." *Journal of General Psychology*, XV (1936), 63-106.
7. Spearman, C. "The Battle Between 'Intuitionists' and 'Psychometrists.'" *British Journal of Psychology*, XXIV (1934), 403-408.
8. Wechsler, D. *The Measurement of Adult Intelligence*. Third Edition. Baltimore: Williams and Wilkins, 1944.
9. Wells, F. L. *Mental Tests in Clinical Practice*. Yonkers-on-the-Hudson: World Book Company, 1927.

COUNSELING STUDENTS TOWARD SCHOLASTIC ADJUSTMENT

STANLEY A. STROMSWOLD

C. GILBERT WRENN

University of Minnesota

A FIRST consideration in discussing scholastic adjustment is a rigorous definition of the term. In this paper, scholastic adjustment will be defined in terms of the characteristics exhibited by a well-adjusted student, rather than in terms of scholastic grades alone. It will be necessary, therefore, to consider the characteristics of those students who are making satisfactory progress in the direction of attaining major educational objectives.

The idea of scholastic adjustment may be enlarged by suggesting that the well-adjusted student possesses the six characteristics listed below.¹ First, he has a high degree of intrinsic interest in the subject matter he is studying. Second, he has a positive attitude toward the more detailed requirements of his particular curriculum. This means a frame of mind which will result in a wholehearted acceptance of the challenge of these requirements. Third, he accepts a realistic evaluation of his own aptitudes and personality. Fourth, he has the ability to concentrate for a reasonable length of time on day-to-day tasks. Fifth, he exhibits stability in his goals and in his emotional life. Sixth, he is able to enjoy life in many areas.

This concept of scholastic adjustment is perhaps not as universally accepted and practiced as the concept expressed in terms of scholastic grades. The passing-grade idea of adjustment implies achievement at a specified level in one area,

¹ This idea of adjustment draws upon the psychological concept of the well-adjusted person. For a fuller discussion of normal adjustment, see item 9 in the References.

usually a narrow subject-matter area, as the criterion of adjustment. While this has some merit in professional education by protecting society from incompetent practitioners, it does not constitute a broad enough basis for the counseling of students in general. There are many students whose subject-matter achievement is safely above the threshold of a C average but who are poorly adjusted. These students represent an exceptionally good opportunity for the counselor to promote a fuller development of the nation's human resources. Our concept of scholastic adjustment implies that every student should be brought to accept the responsibility for attaining his optimum proficiency in various areas of living. A democracy needs the maximum development of its ablest as well as of its average citizens.

II.

The foregoing concept of scholastic adjustment implies that the counselor or adviser must view the scholastic problem in the context of the personality of the client in diagnosing and in developing therapeutic measures. For therapy to be successful in modifying behavior, that is, to promote better adjustment, it will be necessary for the clinician to view the client longitudinally as well as to deal with the immediate symptoms that he presents. This will enable the counselor to comprehend the manner in which the symptoms developed. The etiology of the symptom to a large degree will suggest the depth and extent of the psychotherapy necessary. This, of course, will also be influenced by the unique pattern of each individual case. In general, one might expect problems of motivation or the incongruency of ability and objectives to require deeper therapy than the mere lack of study skills and habits. The latter is largely a matter of teaching, while the former is a matter of the reorganization of the student's attitudes toward himself, or it may be the release of tensions resulting from conflict. The outcome of such reorganization or release will be more effective behavior and a better-adjusted person.

The clinical interview in which objective measurements play a part is an effective procedure in diagnosis as well as in

therapy. In the counseling interview, the clinician will have to seek out the primary causes of maladjustment while attempting to deal with the immediate attitudes and needs of the client. Often it will be necessary to deal with the immediate secondary symptoms before the primary causes may be uncovered. For example, students frequently present displaced emotional reactions such as antagonism toward professors, for which some release can be found through a permissive situation with free expression of these emotions. Free expression may lead to an awareness of the responsibility of the individual in the situation and to a more realistic approach to the solution of the client's problems.

The next thing to be considered is the attitude which the client takes toward his problems. Here, we frequently find such tendencies as withdrawal with negativism, indifference toward the environment, compensation in other activities, rationalization, and dependency. Whatever the attitude presented, the counselor should have as his goal the freeing of the client from attitudes which prevent him from accepting a realistic evaluation of his problems. This will make it possible for him to make a more direct attack upon immediate problems and needed skills.

III.

As a guide for the counselor it is well to set up several diagnostic constructs or possible hypotheses of need (2). The five constructs herein listed will include the majority of the cases presented on the average college campus.

1. The first diagnostic construct is misclassification. Students included in this category do not have the aptitudes, interests, and other personality factors demanded by the particular curriculum which they are pursuing. Since the student does not have the personal resources demanded by his environment he is precipitated into a persistently non-adjustive state. The symptoms presented by the client, which should result in the diagnosis of misclassification, are fairly evident to the counselor. The client, on the other hand, is usually not aware of the meaning of these symptoms. These symptoms include

aptitude scores which show a small probability of success, failure to meet prescribed standards in spite of strong motivation and effort, considerable anxiety, rationalization, and projection. Most of the difficulty in this type of student arises from his resistance against acceptance of the inappropriate nature of his goals in relation to his aptitudes. A frequent reason for his refusal to accept reality is that it represents to him an attack upon his ego and a threat to his social acceptability. The therapy in this type of situation consists largely in getting the client to see that he can achieve personal satisfaction and social acceptability by pursuing other goals. This will free the client from the need of defense mechanisms and will enable him to make a realistic appraisal of himself and his projected program.

2. The second diagnostic construct is inadequate educational background. The important symptom here is that of academic work being performed at a level considerably below the client's potentialities. This type of client makes up a considerable percentage of that group of students usually referred to as underachievers. Inadequate educational background may be present in only one or several areas. This problem is often met in dealing with the more technical curricula. For example, students admitted to third- or fourth-quarter French on the basis of high-school work sometimes find that their high-school work was of such poor quality that they are not prepared to pursue the more advanced work. Or a student may find it almost impossible to follow his physics course because he is unable to manipulate even the simplest elementary algebra. The therapy for these clients is obviously, by one means or another, that of providing the opportunity to learn the prerequisite information and skills. Many campuses now provide courses in sub-freshman mathematics, sub-freshman English, and other fields to meet these important educational needs.

3. The third diagnostic construct is inadequate command of study habits and study skills. This group, like the second, contributes a considerable percentage of the underachievers. Here, however, the deficiency is not so much in a specific sub-

ject-matter area as in inadequate methods of attacking learning problems. Besides the usual reading and writing skills which are often found deficient, we find that college work demands the formation of a number of new skills, such as taking lecture notes, the use of the library or wide reading on a topic, and efficient techniques for memorization. Here, the therapy is a matter of stimulating the client to learn the skills and to form the habits necessary to meet the challenge of college work. This can be done by providing reading clinics, how-to-study courses, and by counseling. Not only can the counseling interview be used as a means for selecting those students who are most apt to profit from these experiences but it also can be used as a means of bringing about learning readiness. Counseling should lead the client into a realization of the necessity for acquiring more effective study habits and skills. Where such courses and clinics are not available, or when the counselor considers it more prudent, the interview may be used as a means for pointing out to the client more efficient techniques of study. When this method is used, follow-up interviews are very desirable. Effort must be made to see that the learning becomes functional. At these follow-up interviews, the counselor should discuss with the client how well these new procedures are working out in his individual case. At this time modifications to fit the individual needs of the client should be made.²

4. The fourth diagnostic construct is inadequate use of time. (Inadequate use of time may be a symptom of some other disorder as well as a separate diagnostic construct.) This is also responsible for many cases of underachievement. Here, we are confronted with the necessity of leading the client to change his habits. To a certain extent, we should stimulate the student to routinize his daily life to make room for his work and his play. As a preparation for doing this, the student should analyze the nature of the distractions which keep him from working during the time he sets aside for that purpose.

² The use of study manuals is also often effective for these clients. Items 3, 8, 11, and 12 in the References are excellent examples of concise and practical manuals for that purpose.

In most cases, some type of time budget will be desirable. A student should be encouraged to schedule study periods as he schedules his lectures and laboratory work. An interesting observation may be made at this point, in setting up time budgets—students are apt to set up a budget which is not workable because it is too ambitious. A good way to avoid this is to suggest that a student begin by planning his work only a day at a time. Students should be encouraged to make out a schedule of the next day's activities in the evening. After having done this for a week or more, the student may gradually learn how much time must be spent on each lesson and how he can best fit his work periods into his daily life. After following this procedure for one or two weeks, the client can build upon these experiences to set up a time schedule for a week's work. To facilitate habit-formation, it is well for him to attempt to study the same thing at the same time daily. A time schedule built up in this manner has the advantage of meeting individual needs much better than one which is set up according to the ideal schedule for that mythical entity, the average person.

5. A fifth diagnostic construct is interference with study by problems outside of the scholastic realm. The student who is constantly in emotional turmoil because of such common problems as conflict with parents, inability to make proper social adjustments, or financial problems, cannot be expected to do his best. It is now recognized that in compensating for such problems, students may become scholastic overachievers. In harmony with the idea of scholastic adjustment presented in this paper, these students are not, however, the "good" students. In this situation, the counselor, through a permissive atmosphere, should encourage the client to discuss his problems and provide the means for self-acceptance and self-resolution of conflict. Study habits, time scheduling, and reading clinics are apt to prove of little value in such cases. No counselor can afford to restrict his counseling to study problems. Each client presents a unique pattern of characteristics and must be dealt with according to his own felt needs.

Careful attention to the adjustment problems of students is one of the most promising means of increasing the value of

education to society. Not only should it serve to decrease the number of maladjusted and disgruntled students who fail scholastically but it should also perform the more important function of stimulating the able student to perform on a level commensurate with his ability rather than to slide through with a minimum of effort. The loss to society through educational underachievement on the part of capable students is very great. It is axiomatic that a democracy can not rise above the level of its citizenry. The counselor who stimulates these capable individuals to a fuller utilization of their abilities is performing a significant social service.

REFERENCES

1. Bird, Charles and Bird, Dorothy M. *Learning More by Effective Study*. New York: D. Appleton-Century, 1945.
2. Bordin, E. S. "Diagnosis in Counseling and Psychotherapy." *Educational and Psychological Measurement*, VI (1946), 169-184.
3. Cole, Luella and Ferguson, Jessie Mary. *Students' Guide to Efficient Study*, Third Edition. New York: Rinehart and Company, 1946.
4. Crawford, Albert Beecher. *Incentives to Study*. New Haven: Yale University Press, 1929.
5. Crawford, Claude C. *Studying the Major Subjects*. Los Angeles: University of Southern California, 1930.
6. Fenton, Norman. *Mental Hygiene in School Practice*. Stanford University: Stanford University Press, 1943.
7. Fry, Clements C. and Rostow, Edna G. *Mental Health in College*. New York: The Commonwealth Fund, 1942.
8. Gerken, C. d'A. *Study Your Way Through College*. Minneapolis: Perine Book Company, 1943.
9. Masslow and Mittleman. *Principles of Abnormal Psychology*. New York: Harper and Brothers, 1941. Chapt. 5.
10. Williamson, E. G. *How to Counsel Students*. New York: McGraw-Hill Book Company, 1939.
11. Wrenn, C. Gilbert. *Practical Study Aids*. Stanford University: Stanford University Press, 1933.
12. Wrenn, C. Gilbert and Larsen, Robert P. *Studying Effectively*. Stanford University: Stanford University Press, 1941.

A PROCEDURE FOR OVERPRINTING ANSWER SHEETS FOR HAND SCORING WHICH MIGHT BE ADAPTED TO LOCAL SCORING

ARTHUR E. TRAXLER
Educational Records Bureau

IN connection with the spring testing programs of 1946 and 1947 carried on by member schools of the Educational Records Bureau, the responses to the *Cooperative English Test* were recorded by the pupils on separate answer sheets regardless of whether the tests were to be hand scored or machine scored. For all tests which were to be hand scored, the scoring key was overprinted on the answer sheets before the scoring process was begun. Circles connected by short, straight lines were printed around the right answers. The scoring procedure consisted of counting the number of responses which fell within the circles, then counting the number outside the circles, and finally of subtracting from the correct responses a proportion of the wrong responses. On the complete English test, including Tests A, B, and C, the scoring time saved by this procedure, in comparison with the time required for scoring when the answers were written in the booklets, was 6.4 minutes per paper, or approximately 42 per cent.

The overprinting not only saved a significant amount of scoring time, but it also facilitated the use of the test results. When separate answer sheets are scored in the usual way, either by hand or machine, one limitation is that no indication is made on the paper of which questions were answered correctly. Thus, it is difficult for the pupil, teacher, or counselor to study the paper and to determine the questions on which the pupil did well and those on which his performance was poor. The overprinting makes an analysis of this kind comparatively easy, provided, of course, that a booklet containing the questions is available.

The diagnostic and reteaching advantages of overprinting were thought to be sufficiently large to warrant the overprinting of answer sheets scored by machine, as well as those scheduled for hand scoring. In the case of the machine-scored answer sheets, however, the overprinting was done after the scoring was completed.

The overprinting was done on a multilith, number 1250. In order to insure that a sufficient number of copies could be made without the necessity of preparing new plates periodically, the various answer keys were photographed on multex plates. During the overprinting, the answer sheets could be run through the multilith at a speed of about 3,000 copies an hour. However, great care in operation was imperative in order to obtain as nearly perfect registration as possible and to guard against the danger of the answer sheets becoming jammed in the machine and damaged or destroyed.

This experiment suggests some interesting possibilities for the use of overprinting in the local scoring of objective tests. To an increasing extent, test publishers are furnishing answer sheets with their tests, and, in the case of a considerable number of new tests, the use of answer sheets is being made a requirement, since there is no provision for writing the answers in the booklets. Yet, many public schools—even some of the larger school systems—are not yet equipped with scoring machines. The hand scoring of the answer sheets is either a long and tedious clerical job for teachers or counselors or the school has to provide a sizable budget for clerical scoring help. It would seem that the overprinting procedure might save public schools hundreds of hours and thousands of dollars.

Even where the scoring is now provided for on an adequate basis either by machine or hand methods, overprinting might be worth the cost through improving teacher, counselor, and pupil use of the results.

There are, however, certain technical problems in the adaptation of the overprinting procedure to local use. Comparatively few schools are equipped with multiliths. Most schools do have mimeograph machines, but the registration with the mimeograph is not as accurate as it is with the multilith. The mimeograph is often beset with paper-feed troubles with the

resultant crumpling and tearing of several pages before the machine can be stopped. Moreover, the IBM answer sheets for certain tests are too wide to be fed through a standard mimeograph.

On the other hand, there is some evidence that overprinting with the mimeograph is feasible. Some years ago Stenquist¹ reported that the labor of scoring was reduced through the use of the mimeograph in overprinting answer sheets for tests given in the Baltimore Public Schools. An experiment run with the mimeograph at the Educational Records Bureau indicated that the registration was sufficiently accurate on about 99 per cent of the answer sheets.

If there were sufficient demand for the adaptation of answer sheets to the requirements of overprinting, perhaps test publishers could use care to see that special answer sheets prepared for use with their tests were of a width such that a standard mimeograph would take them. Another way in which test publishers might aid schools with this type of scoring would be to furnish their answer keys on stencils ready for use in overprinting.

A question might be raised concerning the possibility that an organization such as the Educational Records Bureau could at a very low charge set up an overprinting service for public schools that desired to receive that type of aid with their scoring and to handle the rest of the scoring procedure locally. The addition of this kind of service to the Bureau's present services would not be possible during the fall and spring testing programs without the purchase of more equipment and the expansion of the machine room staff. This type of commitment might be undertaken if the Bureau could anticipate a steady volume of overprinting work. It is probable, however, that there would be much demand for this service in the early fall and late spring and little demand during the rest of the year. Thus, it is likely that the overhead costs would be prohibitive.

It is believed that it would be desirable for public schools to experiment with the overprinting of answer sheets by means of their own facilities.

¹ J. L. Stenquist. "Experiments with Machine Scoring of Tests." *Baltimore Bulletin of Education*, XIII (1935), 83-85.

SELF-APPRAISAL OF TEST PERFORMANCE AS A VOCATIONAL SELECTION DEVICE¹

GEORGE S KLEIN

The Menninger Clinic

THE present paper will, first, describe a selection device based on a modified level of aspiration technique which was developed in the AAF Aviation Psychology Program during the recent war, and secondly, it will demonstrate its specific applicability and validity in an AAF occupational selection problem: namely, the prediction of success or failure in flight training of several hundred aviation cadets. It is implicitly suggested in the following discussion that the measurement of the appropriateness of self-appraisal of performance, that is, the gap between predictions and actuality, may have two principal applications: it may provide useful adjuvant information in vocational selection programs where job analyses suggest that this aspect of ego functioning can be determinative of success or failure in certain occupations, and it may provide also a means of probing an important facet of ego functioning for clinical diagnostic purposes. Here, however, only its vocational application will be illustrated.

Current personality theories essentially agree that an individual's evaluation of himself—his self-esteem—occupies a central position in the functioning of his ego, guiding the predictions he makes concerning his behavior and influencing the goals he sets for himself. Indeed, the pattern of such self-evaluation is considered by some to reflect reliably certain features of the ego's defensive structure itself (2, 3) and even its intrinsic strength or weakness. By this conception, ego structures are distinguished by characteristic patternings of

¹ This study was conducted as part of the AAF Aviation Psychology Program. The interpretations and opinions expressed are those of the author and do not necessarily reflect the official ones of the AAF Aviation Psychology Program.

self-esteem defenses which are invoked for the ego's protection particularly in situations of challenge or threat. These defenses are reflected in the tendency for the individual either to regard his capacities and achievements realistically or consistently to under- or over-evaluate them. Experimental investigation of this facet of ego functioning ("self-evaluation" or "self-esteem") has been greatly facilitated by the level of aspiration method, developed by Kurt Lewin and his associates, and refined more recently by others. One virtue of the method is that it lends itself to adaptation as a testing instrument with clearly delineated, quantitative indices of anticipated performance, actual performance and final self-evaluation of performance, as well as indices summarizing the relations existing among these three aspects. Accordingly, in developing a selection device the present study utilized one aspect of the level of aspiration technique, namely, the measurement of the discrepancy between self-estimates of performance obtained before and after given tasks, and actual achievement on the tasks themselves.

Underlying the present application of the test to an aircrew selection problem was a hypothesis which emerged from several job descriptions of the pilot's functions: that the quality of appraisal of one's performance is an important attribute of success or failure in pilot flying training. In seeking a measure of this factor in the selection process, it was then further assumed that in a test situation self-evaluation of test performance would be distorted in accordance with the disposition of an individual to exaggerate or depreciate what he is able to do.

Sufficient evidence is at hand in the literature to caution that for the latter assumption to hold good in a testing situation, both test performance and self-evaluations must be "ego-involved," that is, that these must be regarded by the subjects as having some intense personal significance. Previous studies have indicated (4) that it is in only such a charged context that characteristic features of the ego defensive structure are likely to come into play. Otherwise it would be as reasonable to propose that an individual's evaluation of his performance in a test situation will closely parallel his experienced difficulty on the task. Whether or not such evaluations will be independ-

ently governed by personality variables lies in the significance they hold for him, either in that they may, in his eyes, potentially influence some vital change in his life situation and affect his achievement of vital goals, or in some other way represent a threat or challenge with consequences extending beyond the experiment proper. This atmosphere of ego-involvement pervaded the AAF classification process of which the present test was a part, and the ratings themselves assumed the type of significance described. This characteristic of the testing situation will be more specifically discussed below.

Description of the Test

To measure the pattern of an individual's self-appraisal, a simple rating procedure was developed and administered in connection with the routine psychomotor tests of the aviation cadet classification test battery. Each cadet estimated his performance before and after each of the 6 psychomotor tasks. One estimate was made before practice and one immediately after completion of the psychomotor test. These will be referred to respectively as *pre-performance* rating and *post-performance* rating in subsequent discussion. On 4 tests a rating was obtainable after a practice period as well so that a total of 16 ratings were made by each cadet. The 6 psychomotor tests were as follows:

1. *Aiming Stress Test*: This was a steadiness task in which the cadet was directed to keep a stylus, held in outstretched hand, from touching the side of a narrow aperture, while verbal stimuli designed to "rattle" him were shouted by the examiner. The measure of unsteadiness was the amount of time the stylus made contact with the sides of the aperture as recorded automatically by electric clocks wired into the apparatus.

2. *Rotary Pursuit Test*: In this test the cadet was required to manipulate a stylus in such a manner as to maintain contact between the point of the stylus and a target set on a rotating disk. Efficiency of performance was determined by measuring the time during which the cadet maintained a direct contact between the stylus and the target.

3. *Complex Coordination Test*: This was a coordination test in which the candidate had to manipulate simulated air-

plane controls—a stick and a rudder bar—in order to match three pairs of lights presented on a panel before him. The measure of efficiency was the number of matchings accomplished within the 15-minute test period.

4. *Two-hand Coordination Test:* This test required an individual to coordinate the movements of both hands in response to a target moving at varying rates of speed along an irregular pathway. Contact had to be maintained between a metal button controlled by two handles which the cadet manipulated and the moving metal target. The measure of efficiency was the amount of time such contact was maintained as recorded by the electric scoring clocks during the series of trials.

5. *Discrimination Reaction Time:* This test was designed to measure the speed of selective response to the spatial relationships existing among 4 visual stimuli. The test required that the cadet make 4 different manual reactions, depending upon which of the 4 positions in space a red and a green light occupied with respect to each other. The cumulative time taken by the candidate to make 80 reactions (4 sets of 20 each) was recorded on the electric clock and constituted his test score.

6. *Finger Dexterity Test:* In this test the speed with which the cadet reversed and inserted pegs into several rows of a pegboard was determined. Efficiency was measured in terms of the number of pegs appropriately placed during a given test trial.

Each of the above tests was given to a group of 4 subjects at a time. For 4 tests the standard testing period consisted of an instruction period, practice trials, and test trials interspersed with 30-second rest periods. On 2 tests, Rotary Pursuit and Two-Hand Coordination, there was no practice period. The testing time for each task was approximately 15 minutes. The sequence of tests was deliberately randomized within the group studied. Performance scores of all tests were rendered comparable by conversion to a standard distribution according to the conventional practice followed in the AAF classification program.

The Rating Scale and Motivational Context: A 10-point scale was devised for expressing self-ratings of performance on

the psychomotor tests. Instructions regarding its use on each task were read to the cadets before they began the series. In this scale, running from "very poor" to "very good," the middle position, "average," was divided, thus forcing ratings to fall either into above-average or below-average parts of the scale. It was believed that in thus forcing the cadet to commit himself, the procedure would provide a measure of a significant personality tendency.

In any level-of-aspiration study, the type of instruction given and the motivational context or atmosphere of "ego-involvement" in which ratings are made are both crucial: they invariably affect the general direction of the ratings of the group of subjects and in part determine the range within which individual differences occur.

The instructions stressed the importance of a "realistic" attitude in predicting performance. This emphasis is illustrated by the following excerpts from the instructions given the cadets.

Very frequently, after taking a test, cadets remark. "I did well on this test," or "I think I did very poorly." Often, even before they take a test, they can size up whether they will do well or poorly. Now before taking a test you will tell us how well you think you *will do*, and after completing each test, you will tell how well you think you *have done*. The purpose is to see how accurately you can estimate your performance. We would like to find men who are realistic, who know themselves well and can judge accurately and frankly their own abilities and performance. By "*realistic*" we mean, if you think you will be "very good," *frankly* say "very good"; if you think you will do very poorly, *frankly* say "very poorly"

A final paragraph admonished the cadets as follows:

You may have heard about these ratings from other cadets. You may have heard, for example, that a certain rating is always "best" or "correct" regardless of what you think you will do or have done. But no one can know whether a high rating or a low rating is better. No particular estimate is always "correct" or "incorrect." Remember, what counts is how close your estimates are to what you actually do on the test. *Be frank*; if you think your score will be high, or if you think it will be low, or if you think it will be somewhere in the middle, say so by marking in the appropriate box.

The instructions further emphasized to the cadet that accurate estimates even of poor performance, i.e., where the discrepancy between prediction and a low score was minimal, would gain him more credit in his final classification score than would high performance with greatly discrepant ratings. This stress on "realism" perhaps tended the group toward conservatism in making ratings, particularly since they were given no objective signs in the form of scores, to indicate how well they were actually doing. However, the instructions had the more important effect of underlining the *independent* significance of the *ratings* in the cadet's final classification standing. In this way, the ratings became for him as crucially important as the task performance proper. Considering also the general atmosphere of high motivation, intense competitiveness and eagerness to qualify existing during this period of the war, we may reasonably infer that all ratings were highly "ego-involved," that they were gathered in a significant situation in response to which the system of self-esteem or self-evaluation defenses of the ego could come into play, and influence the trend of each individual's ratings.

We emphasize this feature of "personal significance" in the testing situation because it is one rarely if ever achieved in level-of-aspiration experiments. Generally these studies have also overlooked the more specific and equally crucial question of how much significance the subjects impute to the ratings themselves, quite apart from the tasks. The mistake is usually made of assuming that because the tasks performed by the subject had personal significance for him (i.e., were ego-involved), that by this fact the ratings or aspiration judgments based on these tasks had a similar significance. However, one frequently finds that where the tasks were ego-involved, the aspiration ratings were not, the subject being aware of the purely experimental character of the ratings and hence of their fundamental irrelevance to the tasks (and, therefore, to himself). In few level-of-aspiration experiments have self-ratings themselves seemed to carry to the subject the significant implications for achieving major life goals that were described in the present study.

Several other departures from the usual level-of-aspiration procedures may be noted: (1) Although the tasks were unfamiliar to the subject he was asked to make pre-practice as well as post-practice ratings; (2) The subject was never given any objective information, actual or "faked," concerning his performance; (3) The subject made his predictions and estimates in qualitative terms, rather than in terms of a quantitative score specific to each task.

A final word concerning the motivational climate in which the study was conducted. The research was carried out during the height of the war, in a period when the prestige connected with flight training and AAF officer-status was especially great and the consequences of failure, especially with respect to the type of combat assignment, particularly drastic. Aside from the popular allure and glamour associated with being an Air Forces pilot, cadets were influenced by popular conceptions concerning the hazards of combat and living conditions in other branches of the service, particularly the infantry. Since popular rumor held that it was usually to these services that cadets who failed at the classification centers were recruited, these conceptions probably further intensified the desire to succeed which was already present and also indirectly contributed to the personal significance that the testing situation acquired.

Validation Indices: The difference-score principle was employed for expressing the relationship between the expressed level of ability and actual performance. In the present instance, the difference scores were measures of discrepancy between expected level of performance ("How well I *will* do" or "How well I have done") and actual achievement (standard score on the test). To permit this comparison, self-ratings were translated into standard score units comparable to those used for the psychomotor tests. The score representing actual achievement was then subtracted from the corresponding estimate of performance to provide a difference score.

The difference scores made by the cadets on each of the tests were transmuted to a 9-point scale, a scale value of 1 indicating extreme underestimate of performance and a scale

value of 9 indicating extreme overestimate. This score has a distribution function with a mean of 5 and sigma of 2. Only the measures of discrepancy based on rating 1 (*pre-performance*) and rating 3 (*post-performance*) were selected for validation.² Each cadet had a total of 12 such scores, two for each test.

The rating scale was administered to 606 cadets taking psychomotor tests. The validation group consisted of those numbers of this population who were immediately assigned to pilot training, in all 243 cadets who eventually succeeded in elementary flight training, and 29 eliminees from this stage of training. The validation criterion was success or failure. A total of 229 of the 606 cadets were assigned to later pilot training classes. The remaining 105 of the original group comprised small numbers assigned to navigator and bombardier training and also those who were eliminated from cadet status at the classification center because their aptitude scores were too low to qualify them for any aircrew specialty.³

Results

Biserial correlations with pass-fail data were computed for each of the difference scores. Since the standard score on the psychomotor test contributes in part to the size of the difference score, it was considered desirable to hold constant (i.e., partial out) this factor in estimating the validities of the difference scores and their independent contribution to the classification test battery. Thus, the correlations of the pre- and post-performance difference scores with standard performance scores were obtained as well as other correlational data relevant to the computation of the partialled validity coefficients of the difference scores.

The unpartialled and partialled correlation coefficients are presented in Table 1 along with the contribution of each difference score to the battery. The additions to the multiple R

² It will be recalled that on four tests predictions of performance were also made immediately after the practice period, but these are not treated in the present study.

³ A second validation study was subsequently conducted on another much larger group but the results are as yet unavailable.

are based on the partial r 's. It will be seen that 7 of the 12 difference scores validate highly enough to add .01 to .05 to a battery validating .57 on the population employed. It should be noted that the present coefficients represent minimum validities, since no correction was made for the somewhat restricted range of the validation sample which resulted from

TABLE 1

*Correlations of 12 Difference Scores and a Combined Weighted Score With Pilot Classification Scores and Pass-Fail in Primary Flying School ($N_p=243$, $N_o=29$)**

		Correlations with Standard Scores Varying		Correlations with Standard Scores Held Constant		Contribution to Validity of December 1943 Battery†
		r_{14}	r_{b1813}	$r_{14\ 2}$	$r_{b1813\ 2}$	
Rotary Pursuit	Pre-P.	-.22	-.30	-.09	-.24	.04
	Post-P.	-.13	-.28	+ .02	-.22	.05
2-Hand Coord.	Pre-P.	-.33	-.32	+ .10	-.08	.02
	Post-P.	-.37	-.37	-.04	-.19	.03
Discrim. RT	Pre-P.	-.25	-.15	-.08	-.10	.00
	Post-P.	-.33	-.30	-.22	-.27	.02
Aiming Stress	Pre-P.	-.24	-.19	-.04	-.14	.01
	Post-P.	-.25	-.14	-.11	-.08	.00
Finger Dex	Pre-P.	-.18	-.15	-.10	-.11	.00
	Post-P.	-.23	-.23	-.17	-.20	.01
Complex Coord.	Pre-P.	-.45	-.33	-.09	-.02	.00
	Post-P.	-.43	-.33	-.09	-.05	.00
Combined Weighted Diff. Score		-.36	-.4506

* Code. Difference Score 1
 Standard Score 2
 Criterion (Pass-Fail) 3
 Combined Classification Battery Score . . . 4

† Validity of battery. .57.

the elimination of cadets at the classification center. It will be recalled that initial selection occurred in terms of the psychomotor test scores. Although there were not many such eliminees represented in the original tested population of 606, this selective factor probably resulted in some reduction in the variances of the difference score indices. Consequently, it is possible that a range correction which would allow for the absence of this low score group from the validation sample would

result in an increase in the estimated validities of the difference scores.⁴

A combined weighted score based on the difference scores was also validated on the same population. This combined score was determined by assigning weights to the various difference scores according to the size of their respective contributions to the battery. For example, the difference score based on the post-performance rating (III) on Rotary Pursuit was assigned a weight of 5 since it yielded a contribution of .05 to the battery. The formula for obtaining the combined score was as follows:

$$\text{Combined score} = 5 \times (\text{Rot Pur: III}) + 4 \times (\text{Rot Pur: I}) + 3 \times (\text{2-hand Coord: III}) + 2 \times (\text{2-hand Coord: I}) + 1 \times (\text{Fing Dex: III}) + 1 \times (\text{Aim Stress: I}) + 1 \times (\text{Disc. R. T: III}).$$

The biserial correlation coefficient of the combined weighted score and its correlation with the combined classification test scores are also given in Table 1. It will be seen that the combined score adds .06 to the multiple *R* of the battery.

In general, the negative direction of the coefficients indicates that with standard score held constant, cadets who overestimate their performance are more likely to fail in primary school than those who underestimate their performance.

Partly as a check on the partial correlations and also for the purpose of studying the degree of linearity of the difference score relationship to pass-fail data, it was decided to employ another method of holding standard scores constant and to observe the relationship of difference scores to graduation or elimination. This method involved obtaining the frequency of graduates and eliminees for each of the difference-score levels in each of the standard-score groups. For this the major breakdown was standard-score; standard scores were grouped regardless of test. First, the standard scores of all the psychomotor tests were tabulated into groups of five; viz., 13-17, 18-22, and so forth, up to 83-87, 10 groups in all. Then, the

⁴ It should be noted also that battery and test validities were severely restricted by the fairly low reliability of the pass-fail criterion in the primary schools at the time this study was conducted.

actual per cent of graduates was computed for each of the 10 standard-score groups. Each of the 10 groups contained enough cases to make the results statistically reliable. It was then determined to what degree the various difference scores falling in each one of the standard-score groups were associated with graduation or elimination, and their discriminative capacity was compared with that of the standard-scores. This was done by determining how much the percentage frequency of graduates having each of these difference scores exceeded or was less than the actual per cent graduating in that particular standard-score group. For each difference score, the increment or decrement in per cent of graduates over or below that of the "expected" percentage was obtained in each of the standard-score groupings; these residual percentages were then summed for each difference score. Thus, it was found, for example, that the percentage of men with difference scores of 9 who graduated was much less than the actual percentage of graduates in each of the standard score groupings in which they were found. Following this procedure for each of the nine levels of difference scores in each of the ten standard-score groups, it was possible to determine to what degree a difference score was advantageous so far as graduation was concerned. In this way, also, this procedure provided an indication of the linearity of the difference-score relationship to pass and fail data. The results represent an average relationship based on approximately 3,000 difference scores.

Figure I summarizes the relationship of each of the difference scores to primary school performance, with standard scores held constant. The values of the Y ordinate in Figure I show the total percentage of graduates in excess of or less than the actual percentage of graduates in the standard-score cells in which the difference scores were found. It will be seen that the percentage of graduates with a difference score of 1 was found to exceed by 68 per cent the percentage of graduates in each of the standard-score groups. At the other extreme of the scale, it was found that the percentage of graduates among the 8 and 9 difference scores is much less than the expected percentage of graduates in each of the standard-score groupings

and can be regarded, therefore, as associated with failure in primary school. It thus appears evident that a considerable underestimate of performance is more highly associated with success than any other level of self-estimate; conversely, a con-

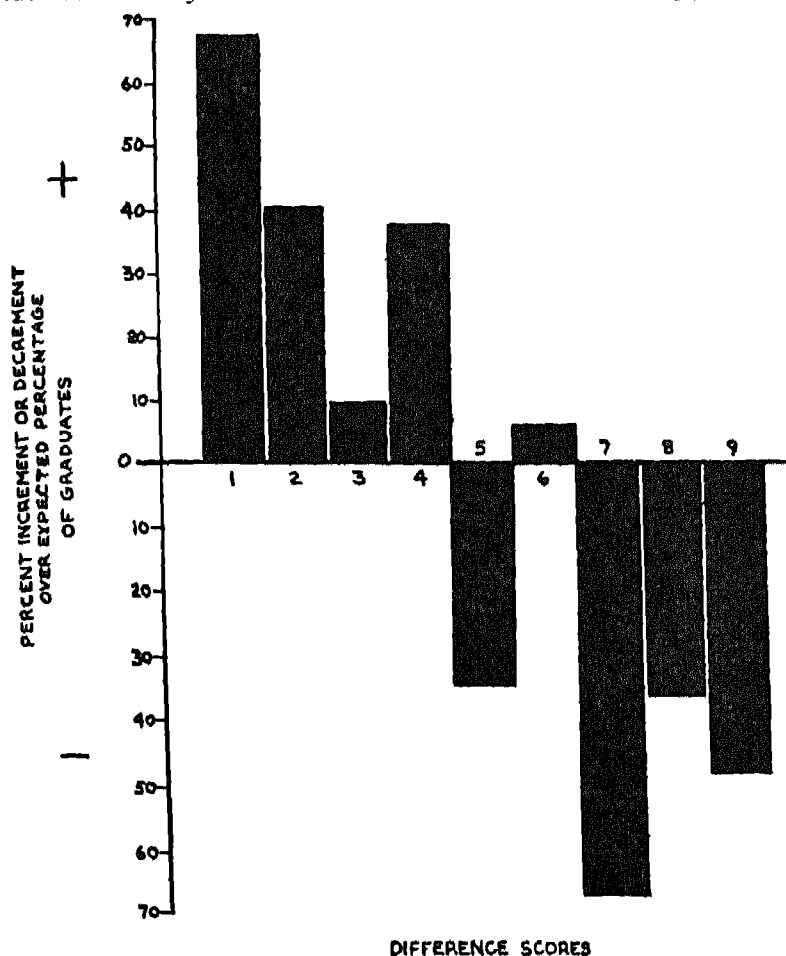


Figure I. Showing the relationship between the various levels of discrepancy scores to primary school success with standard score held constant.

siderable overestimate is closely associated with failure in primary school.

The data in this form substantiate the conclusions derived from the partial correlation coefficients and demonstrate, in

general, the linearity in relationship between difference score and success or failure in primary school. The relatively small number of cases in the final validation group ($N=272$) limits, of course, the conclusion drawn from these data and may also be partly responsible for the reversal of the linear tendency in the middle part of the difference-score scale

Discussion

The finding that underestimates of performance are associated with success is of considerable interest. It may be that such low self-evaluations are really indirect measures of motivation. For example, a man with a wealth of drive and ambition may depreciate his immediate performance if he measures it against his own absolute standards, which are set high. He will rarely, if ever, come up to his own standard no matter how well he does. In the case of the overestimates of performance we may be seeing in operation a failure in self-insight which is detrimental to success.

The results may be viewed also from the somewhat different vantage point of total ego functioning. This would relate the individual's self-evaluations to the self-esteem defenses that are invoked in situations of threat and challenge (2, 3). From this point of view, overly optimistic estimates of one's achievement may be a useful device to help protect the ego from contemplated effects of failure, especially if kept resolutely high despite poor performance. This would be the case in certain forms of neurotically disturbed egos which particularly require such inordinately rigid defensiveness; egos relatively free from the thwarting effects of neurotic conflict presumably have no such need. This view is expressed by Erickson (1): "A strong ego secure in its identity does not need, and in fact is immune to any attempt at artificial inflation. Its tendency is toward the testing of what feels real; the mastery of that which works; the understanding of that which proves necessary."

It is reasonable to suppose that within certain dynamic settings consistent *underestimates* of performance may also reflect a self-esteem defense. It appears unlikely, however, that the underestimate tendency of the successful group in the pres-

ent study can be so interpreted. Such an interpretation would specifically imply that a neurotic quality which expresses itself in a need to "play down" one's efforts is conducive to successful achievement as a flier. In the present instance, therefore, it appears more reasonable to refer this trend within the successful group to the conservative bias induced by the rating instructions, rather than primarily to a characterological source. It will be recalled that the implicit emphasis on caution probably had some tempering effect, to a varying extent, on all individuals tested. In such a situation, a "realistic" individual would characteristically tend toward conservatism or underestimates in his predictions, the direction of the successful group. In the failure-group, on the other hand, pressing defensive needs might well have overcome the conservative emphasis of the instructions. Presumably, were the instructions such that a "realistic" attitude could be expected to show itself consistently in zero-discrepancy scores, one would find that the degree of overestimation characteristic of the failure-group would be much greater, and the degree of underestimation in the successful group would be much less, than was the case in the present study.

The foregoing discussion provides little basis for explaining why the difference scores varied so widely in validity or why some showed no validity at all. It may simply be that some of the tasks afforded no stable comparison or "anchoring" basis according to which an individual could relate his performance to that of other individuals or estimate his own progress during the test. It may be, however, that some tasks tend to call into play self-esteem defenses more readily than do others. It would seem from the results, also, that first and last ratings do not consistently bear the same significance from task to task with respect to ego factors or, more specifically, with respect to their relationship to self-esteem defenses.

Unfortunately, the data gathered were insufficient to clarify the dynamics underlying the trends described. Nevertheless, one generalization from the present findings is worth special emphasis because of its practical implications: that in at least one type of vocational adjustment, the quality of appraisal of

one's performance is an attribute of success or failure, and that the technique described affords a means for measuring it. This suggests that vocational testing programs may gainfully measure this factor for predicting success in other occupations as well, especially where job analyses appear to indicate its relevance. This possibility offers a pattern for future research.

Summary

1. The present study describes a selection device based on a modified level-of-aspiration technique and its application to the prediction of flying school success or failure in a group of 272 aviation cadets.

2. Measures of discrepancy were obtained for each cadet between self-estimates of performance made before, during, and after each of 6 psychomotor tests, and the standard scores he obtained on the test proper. The difference scores were expressed on a scale of 1 to 9. Twelve difference scores, based on the pre-performance and the post-performance ratings, were validated.

3. It was found that 7 of the 12 difference scores validated highly enough to contribute from .01 to .05 to a battery validating .57 on the population employed. These additions to the multiple R were based on partial r 's, with performance scores on the tests themselves held constant.

4. A combined weighted score, determined by assigning weights to the various difference scores according to the size of their respective contributions to the battery, was also validated on the same population. This was found to yield a contribution of .06 to the multiple R of the battery.

5. In general, the direction of the relationship of self-estimates to performance in primary-school flying training appears to indicate that with standard score on psychomotor tests held constant, cadets who overestimate their performance are more likely to fail in flying training than those who underestimate their performance.

6. It is felt that the promising nature of these results warrants a wider application of the technique to problems of vocational selection in addition to its possible applicability to clinical diagnostic problems.

REFERENCES

1. Erikson, E. Homburger. "Ego Development and Historical Change." *The Psychoanalytic Study of the Child*, Vol. II. International Universities Press, 1946. Pp. 359-396.
2. Frank, J. D. "Recent Studies of the Level of Aspiration." *Psychological Bulletin*, XXXVIII (1941), 193-226.
3. Holt, R. R. "Level of Aspiration: Ambition a Defense?" *Journal of Experimental Psychology*, XXXVI (1946), 398-416.
4. Klein, G. S. and Schoenfeld, N. "The Influence of Ego-Involvement on Confidence." *Journal of Abnormal and Social Psychology* XXXVI (1941), 249-258.

A FACTOR ANALYSIS OF THE BERNREUTER PERSONALITY INVENTORY

GLENN C MARTIN
Los Angeles City College

THE *Bernreuter Personality Inventory* is still one of the prestigious instruments of its type in the personality testing field, even since the advent of a number of others with better empirical validation, such as the *Minnesota Multiphasic Personality Inventory*, or with better statistical ground-work, such as the *Guilford Personality Inventory*. In our efforts at the Counseling Service to make the best use of the best available tests, we were forced to consider the Bernreuter as a candidate.

We thought of it rather distantly because of the mixed evidence of its value—for instance, a lack of agreement between the Inventory scores and the actual behavior records of maladjustment. Feder and Baer (3) point out that

Personality problems, arising only in the face of certain specific circumstances, were frequently obscured by the generalized type of questions contained in the Inventory. Cases which should be recommended for psychiatric treatment in terms of the norms, actually manifested problems of less seriousness than other cases given satisfactory adjustment ratings.

Further, its failure to measure socially evident characteristics is criticized by Hanawalt and Richardson (4) who feel that self-sufficiency is distinguishable from withdrawal, and sociability (as evidence of social dependency) from gregariousness (as evidence of mature self-confidence with others), and that the Bernreuter blurs these distinctions.

Usually the reliability of such inventories is determined by a measure of internal consistency, often yielding coefficients as high as .85 for the Bernreuter scales (10). These results, however, are much more favorable than test-retest measurements, reported by Sartain (8) to lie between .523 and .686.

Of course, the consistency of repeated measurements is partly a function of the measured variable as well as the reliability of the measuring instrument, so we cannot assume such a serious fault in the Bernreuter until we know more about the evanescence or stability of personality qualities.

Another slighter difficulty lies in the method of indicating responses. Bernreuter uses the question form, requiring "Yes," "No," or "?" answers. Feder and Baer (3) point out that the rather general statements cannot be answered in such unqualified terms and still leave the feeling that an accurate personal description results. Of course, it should be noted that logically it is quite possible to get a quite satisfactory measure without leaving that feeling. A person unacquainted with X-ray technique might feel disappointed with a moment's exposure to nothing visible without in the least interfering with the picture. This is somewhat comparable to the virtue of face validity in an aptitude test—more desirable with it but not invalidated without it. This whole issue will be enlightened, I am sure, by a current study by Kimber on Guilford's *Inventory of Factors STDQR* concerning the relative merits of statement and question items, and of two, three, four, and five categories for responses, as well as feelings of acceptability attaching to the different forms.

A variant of this aspect of the Inventory is the problem of weighting responses. Bernreuter uses weights from $+8$ through 0 to -8 , the signs roughly indicative of the bipolar characteristic of the scales. McClelland (6) suggests reducing this 17 point continuum to a 7 point range, finding that the simplified scores correlate higher with the original scores than the split-halves of the latter do with each other. Bennett (1) goes even further, reducing the Bernreuter weights to the values, $+1$ and $+2$.

We turn now to another more fundamental difficulty. Experimental workers have found that the measurements by the various scales are not as separable as their titles. St. Clair and Seegers (7) state "The F1 and B1 scores seem to measure nearly identical traits." Feder and Baer (3) assert that the relationship "between and neurotic (B1), introversion (B3) and sub-

missiveness (B4) and self-consciousness (F1) suggest widespread overlapping in these areas. Introversion (B3) and self-consciousness (F1) are highly related as are extraversion (B3) and self-confidence (F1), indicating further overlapping. Dominant traits seem to be associated with self-confidence to a high degree." St. Clair and Seegers (7) found that B3 and B1 are nearly identical, but otherwise found evidence for three different trait measurements: neurotic tendency B1, self-sufficiency (B2), and dominance (B4). These studies and many others reviewed by Super (9) point to the same conclusion that Flanagan arrived at 13 years ago: namely, that the Bernreuter scale titles are to a large extent fictitious, measuring mainly two discrete entities in personality structure. Flanagan used the Hotelling method where "the resulting descriptions of the n variables are in terms of n common factors" (5), and which assumes communalities of unity. His results are shown in Table 1.

TABLE 1

Scale	I	II	III-IV
Neurotic Tendency887	.228	$\sqrt{-.023}$
Self-Sufficiency	-.594	.648	.167
Intro-Extroversion858	.321	.084
Dominance-Submission	-.833	.112	.358

On the basis of this evidence Flanagan constructed two additional scales, contending that they could well be substituted for the original four.

Bennett (2), in addition to simplifying the weighting scheme, advocates using the two F scales as axes of measurement and then, on tables based on regression equations, looking up the most probable scores on the four B scales. He discards the descriptive value of the F scales, however, "since these do not aid significantly in understanding an individual." This confirms Super's feeling that the Bernreuter scales have been retained because of their colorful and familiar verbiage; while the Flanagan substitutes have been largely neglected from a lack of appreciation of their clearer rationale and better statistical basis.

The purpose of this study was to analyze a matrix which includes the Flanagan scales with a method that has come to be much more widely accepted than the Hotelling; namely, the Thurstone centroid method. Verification of Flanagan's results would reduce the probability that statistical artifacts were being created, and at the same time give greater confidence in an abbreviated treatment of the Inventory. The raw material is the matrix given in the 1935 Bernreuter manual.

The rotated factor loadings are shown in Table 2.

TABLE 2

Scale	Factor Loadings			Variance			Communality h^2	Uniqueness U
	I	II	III	I ²	II ²	III ²		
B1-N	.998	.083	.045	.996	.007	.002	1.0	0
B2-S	-.412	.747	.094	.170	.558	.009	.737	.263
B3-I	.959	.106	.234	.920	.011	.055	.986	.014
B4-D	-.829	.127	.411	.687	.016	.169	.872	.128
F1-C	.995	.136	-.081	.990	.019	.007	1.0	0
F2-S	.198	.820	.301	.039	.672	.091	.802	.198

The first point to note is the agreement with Flanagan's factor loadings. He used only the original 4 "B" scales, so that the comparison is not complete, but is still striking.

We have good reason to hold that factor I is the most important in scales B1-N, B3-I, B4-D, and F1-C. If we accept the validity of the Bernreuter scale descriptions, we must accept, too, the conclusion that whatever causes high scores on the neurotic tendency measure must also cause or be highly correlated with the cause, of similarly high scores in introversion and self-consciousness. The negative loadings on B2-S and B4-D would indicate that this factor is inversely related to the causes of social dominance, and to a lesser degree, to self-sufficiency. That the two latter variables are not identical is shown by two bits of evidence: (1) factor II loading on B2-S is heavy, but vanishing on B4-D, (2) some third factor has an appreciable loading on B4-D but vanishing on B2-S. However, there may be some overlapping, since the variance of these two scales is not fully accounted for by the three factors shown. Because of the location of its heavy loadings on B1-N, B3-I

and F1-C, we might tentatively entitle this factor I "Neurotic Self-Consciousness."

Factor II has the heaviest loading in a non-social or socially independent characteristic, with a second on self-sufficiency. The picture is not quite so clear here, but there is probably reason for calling factor II "Social Self-Sufficiency."

It would be dangerous even to suggest any identification for factor III because of the low or vanishing loadings throughout the six scales. (Parenthetically, the writer suggests that these factor titles be held very tentatively, pending some better validation than correlation with "previously validated tests," upon which Bernreuter's claim for validity of the Inventory rests [see *Manual*, page 4].)

After evaluating the research evidence accumulated to 1942, Super (9) comes to a conclusion quite in line with the above factor definitions.

The F1-C continuum appears to consist of good contact with the environment as opposed to poor; the F2-S, of wholesome liking for contact with others in the environment as opposed to wholesome liking for freedom from such contacts. The F scales are thus concerned, the one with type of contact, the other with need for contact, whereas the B scales confuse type and need.

Turning to the consideration of variance, we can legitimately average the values under I and say that factor I accounts for 63% of the total variance of the 6 scales. B1-N and F1-C can be said to be completely composed of factor I with B3-I approaching this limit. Doing the same for II we get 21%; for III, 6%; for U, 10%. Thus factors I and II include 63% plus 21%, or 84% of the total variance. Adding factor III, we get 90%, leaving 10% for additional factors that might be present in B2-S, B4-D, and F2-S, for uniqueness not ascribable to any common factors, and for variance due to errors of measurement. Omitting column U because of the undifferentiated nature of its variance and column III because of the difficulty in identifying that factor, we see that still 84% of the effective measurement of the Inventory can be described in terms of factors I and II.

There are four considerations in the choice of scales to represent the two factors.

They are, in order of importance: (1) purity, (2) independence, (3) reliability, (4) rationale.

1. The comparison of loadings in Table 2 leaves no doubt that F2-6 is a much purer measure of factor II than any other scale.

2. Independence refers to a relative lack of correlation of two variables with respect to a third, so that this criterion does not apply here.

3. Reliability of scales related to factor II (from the Bernreuter *Manual*):

B2-S	.92 ($\pm .018$)*	.85 ($\pm .024$)	
F2-S			.78 ($\pm .039$)

* Standard errors added by the writer.

While these results are not strictly comparable there is some indication that B2-S would offer some advantage in greater reliability. It should not be objected that test-retest reliabilities might be quite different, since we are interested in the single cross-sectional measurement.

4. Rationale is placed as least important on the grounds that the empirical evidence has prior claim to consideration. If the correlation of the length of one's great toe nail with one's chances of getting to heaven can be demonstrated, I will accept the prediction, regardless of the rationale. On the other hand, if the other three considerations have failed to differentiate, then that scale which is supported with the most adequate explanation, with all due respect to Occam's razor, would be the most acceptable. As Super has pointed out, the F2-S scale has a clearer rationale, so deserves selection on a second count.

Referring now to scales correlated with factor I:

1. For purity, we can look back at Table 2. B3-I shows a distinctly lower factor loading and variance for factor I than the two other scales, leaving B1-N and F1-C in the field. But here the difference is too slight to warrant a choice.

2. Which of the two scales B1-N and F1-C should be selected depends on the degree of independence from F2-S. This will be determined of course, by the inter correlations. Looking back at the original matrix, we find the following:

B1-N	and	F2-S	.32
F1-C	and	F2-S	.11

3. The reliability coefficients for the scales correlated with factor I follow:

B1-N	.91 ($\pm .02$)	88 ($\pm .02$)	
B3-I	.89 ($\pm .025$)	.85 ($\pm .025$)	
F1-C			.86 ($\pm .027$)

Since these coefficients were obtained from three different populations they are not directly comparable, but it can be seen that the B1-N scale has only a questionable margin of superiority over B3-I and B3-I over F1-C. Other things being equal, then the most reliable measure would be the best choice. However, other things are not equal.

4. On the matter of rationale, Super's view is again offered as grounds for preferring F1-C, giving it a second point of preference over its competitors.

With reference to the mechanics of handling the Inventory, we facilitate our scoring by having the examinee mark a separate one-page answer form susceptible of complete scoring with four stencils (two positive and two negative). We use the Bernreuter weights, not because of objections to simplified techniques, but because that gives us direct access to the Bernreuter norm data. This method retains practically the entire measuring value of the Inventory, preserves its maximum discrimination, and brings scoring to a practical compass. Because of the two-factor manifold, graphing centile scores becomes easy in a two-dimensional space. The orthogonal relationship of the reference axes in the Thurstone method makes it proper to use rectangular scales in the graph, so we prepared a form using equal standard score intervals for the presentation of test results.

Conclusions

1. It should not be assumed that because the Bernreuter Inventory fails to distinguish self-consciousness from introversion, for instance, that other measures, such as the Guilford, might not.

2. The success in this study of separating factor I from factor II on orthogonal axes fails to support Guilford's suggestion that oblique axes may be characteristic of analyses of personality factors.

3. This study offers confirmation of the Flanagan analysis, and suggests that most of the measuring effectiveness of the Inventory can be tapped with the F1 and F2 scales.

REFERENCES

1. Bennett, George K. *Directions for Using Simplified Scoring System with the Bernreuter Personality Inventory*. New York: The Psychological Corporation
2. Bennett, George K. "Simplified Scoring System with the Bernreuter Personality Inventory." *Journal of Applied Psychology*, XXII (1938), 390-394.
3. Feder, D. D. and Baer, Opal L. "A Comparison of Test Records and Clinical Evaluations of Personality Adjustments." *Journal of Educational Psychology*, XXXII (1941), 133-144.
4. Hanawalt, Nelson and Richardson, Helen. "Leadership as Related to the Bernreuter Personality Measures." *Journal of Applied Psychology*, XXVII (1944), 397-411.
5. Holzinger, Karl J. and Hauman, Harry H. *Factor Analysis—A Synthesis of Factorial Methods*. Chicago: University of Chicago Press, 1941.
6. McClelland, David. "Simplified Scoring of the Bernreuter Personality Inventory." *Journal of Applied Psychology*, XXVIII (1944), 414-419.
7. St. Clair, Walter and Seegers, J. Conrad. "Certain Aspects of the Validity of the Bernreuter Personality Inventory." *Journal of Educational Psychology*, XXIX (1938), 301-311.
8. Sartain, A. Q. "A Study of Bernreuter Personality Inventory Scores Made by Candidates for Supervisory Positions in an Aircraft Factory." *Journal of Social Psychology*, XXIV (1946), 255-259.
9. Super, Donald E. "The Bernreuter Personality Inventory—A Review of Research." *Psychological Bulletin*, XXXIX (1942), 94-125.
10. Traxler, Arthur E. "Measurement in the Field of Personality." *Education*, LVI (1946), 424-430.

PUPIL PERFORMANCE ON THE IOWA EVERY-PUPIL TESTS OF BASIC SKILLS ADMINISTERED AT HALF-YEAR INTERVALS IN THE JUNIOR HIGH SCHOOL

W. C. KVARACEUS and MARY A. LANIGAN

Boston University

THERE is general agreement that a continuous testing program, using different forms of the same test, will result in a more effective testing program. Wood¹ suggests that any testing will have value to a school in direct proportion to the number of times a test can be administered, thereby permitting a continuous study of pupil growth over an extended period of time. Such a program calls for tests that have many forms and a high degree of reliability.

The customary conception of tests, as being essentially snapshot affairs to be given, scored, and acted upon at particular moment and then forgotten, is now thoroughly discredited in all informed circles. . . . Only by an appeal to history can they be made to fit into the program at all. . . . During the last few years has developed a clearer understanding of the basic necessity of measuring growth over a considerable number of years, and a more adequate appreciation of the inescapable need for a systematic use of tests that yield comparable results.

In keeping with this thought, some of the tests that have been published recently provide four or five forms. The *Iowa Every-Pupil Tests of Basic Skills*² is an example of a widely used series of tests whose five forms permit a continuous testing of achievement on the same scale.

In the Arlington, Vermont, elementary schools all twenty-

¹ Ben D. Wood *Basic Considerations in Educational Testing* Published by the Committee on Educational Testing, May, 1933 p. 4.

² *Iowa Every-Pupil Tests of Basic Skills*. Boston: Houghton Mifflin Co

seven pupils completing grade six were tested, using the *Advanced Iowa Every-Pupil Test*, Form N. This same group was then retested with other forms of the same test at half-year intervals. The Iowa tests were selected for use because of their many forms, because of their diagnostic value, and because of their reported reliability.³ At the same time the *Otis Quick Scoring Mental Ability Test*,⁴ Beta form, was administered in 1943 and in 1945. The average of these two scores has been taken as the best measure of the academic aptitude of the pupils discussed in this report.

TABLE 1
Tests Administered to 27 Arlington, Vermont, Pupils at Five Half-Year Intervals Beginning at Grade 6.9

Test	Form	Date	Grade
Otis, Quick Scoring	Beta	Sept. 1943	6.0
Otis, Quick Scoring	Beta	Sept. 1945	8.0
Iowa Every-Pupil	N	June 1943	6.9
Iowa Every-Pupil	O	Jan. 1944	7.5
Iowa Every-Pupil	L	June 1944	7.9
Iowa Every-Pupil	O	Jan. 1945	8.5
Iowa Every-Pupil	M	June 1945	8.9

The mean Otis IQ of this group was 104.35, *SD* 11.41, the standard error of the mean was 0.44. The group is a near-average sample of an entering junior high-school population and represented all of the children in the grade in this school system.

Table 2 summarizes the means and the variability of the group of 27 pupils on each of the five sub-tests in the Iowa Battery, administered at half-year intervals from June, 1943, to June, 1945. This table reveals to what extent the tests reported increments of growth for the skills measured as well as the extent to which pupils tend to spread at each testing period.

Assuming normal growth expectancy of grade increments in the skills measured during this three-year period, the group tested showed regular gains in Reading Comprehension, Vocab-

³ *Iowa, Every-Pupil Test, Supplement to Manual for Interpretation*, Forms N, O, M, pp. 33, 52, 31.

⁴ *Otis Quick Scoring Mental Ability Test*, Beta Form. Yonkers-on-the-Hudson: World Book Company.

ulary, and Work-Study Skills. Increments in Language and Arithmetic are fairly regular except for the January, 1945, testing when regressions were noted. The possibility of such regression scores being obtained on a given group calls for continuous testing if hazards of misinterpretation of such group performances are to be avoided.

TABLE 2

Means and Standard Deviations of 27 Pupils, Grade 6-8 9, Arlington, Vermont, Tested With Iowa Basic Skills Battery Forms O, L, M, N, O, Administered at Half-Term Periods, June, 1943-June, 1945

Test	Mean Grade Score	Variability	
		+ 1 SD	- 1 SD
Reading Compreh. (June '43) . . .	7-1	9-0	5-0
Reading Compreh. (Jan '44)	8-0	9-9	6-1
Reading Compreh. (June '44)	8-5	9-8	6-6
Reading Compreh. (Jan '45) . . .	9-3	10-0*	7-3
Reading Compreh. (June '45) . . .	9-5	10-0*	7-8
Vocabulary (June '43)	6-2	7-5	5-0
Vocabulary (Jan. '44)	7-3	8-8	5-9
Vocabulary (June '44)	7-8	9-0	6-0
Vocabulary (Jan '45)	8-3	9-7	7-0
Vocabulary (June '45)	8-7	10-0*	7-4
Work-Study Sk. (June '43)	6-3	6-9	5-9
Work-Study Sk. (Jan. '43)	6-9	7-9	6-3
Work-Study Sk. (June '44)	8-3	9-4*	7-2
Work-Study Sk. (Jan. '45)	8-8	10-0*	7-6
Work-Study Sk. (June '45)	9-1	9-9	8-3
Language (June '43)	6-9	9-0	5-6
Language (Jan '44)	7-3	9-5	5-8
Language (June '44)	8-7	10-0*	6-7
Language (Jan '45)	8-4	10-0*	6-9
Language (June '45)	8-9	9-9	8-3
Arithmetic (June '43)	6-5	7-1	6-0
Arithmetic (Jan '44)	7-5	8-7	6-6
Arithmetic (June '44)	9-0	10-0	7-6
Arithmetic (Jan '45)	8-7	10-0	7-5
Arithmetic (June '45)	9-2	10-0*	8-0

* Grade score equivalent above grade 10-0 and not available in norms.

The relationship between the average Otis IQ's and the total test performance of the group was also studied at each testing period. Many studies have been made comparing intelligence test results and achievement test results. In Kohn's⁵

⁵H A Kohn "Achievement and Intelligence Examinations Correlated With Each Other and With Teacher's Ranking." *Pedagogical Seminary and Journal of Genetic Psychology*, (1938), 433-437.

investigation a mean correlation between intelligence and achievement was .54. Gates⁶ reported a relationship between school achievement and general intelligence from .47 to .65, with an average of .54. Byrns and Henmon⁷ reported a correlation between achievement and intelligence at the tenth year of school at .396. Another correlation between school achievement and the scores on the Otis Intelligence Tests reported by Ames⁸ was .54. Over a period of three years a coefficient of correlation between achievement and the Otis Intelligence Tests was .52 as reported by Kreuger. (The average IQ of the last reported group was 110.)

TABLE 3
*Correlation between Otis SA, Averaged IQ's and Iowa Basic Skills Sub-Test
Scores of 27 Arlington, Vermont, Pupils, Grade 6-9, at One-
Half-Year Intervals, June, 1943-June, 1945*

Test	Form N June '43	Form O Jan. '44	Form L June '44	Form O Jan '45	Form M June '45	Mean "r"
Reading Compreh70	.81	.75	.85	.79	.78
Vocabulary77	.61	.74	.77	.68	.72
Work-Study75	.72	.52	.79	.87	.73
Language66	.61	.41	.63	.68	.60
Arithmetic27	.88	.57	.66	.56	.59

In the first battery of tests administered in June of 1943, a correlation between the averaged Otis Intelligence scores and Reading Comprehension was .70; with Vocabulary, .77; Otis IQ's and Work-Study Skills, .75; with Language, .66; and a correlation of .27 with Arithmetic and the averaged Otis scores.

The second set of tests, Form O, administered in January, 1944, maintained a consistently higher correlation than those previously reported. Similar results can be noted in Test Form L, administered in June, 1944, Test Form O, administered in January, 1945, and Test Form M, administered in June, 1945, with three exceptions, maintained a consistently higher

⁶ A. I. Gates. "The Correlation of Achievement in School Subjects With Intelligence Tests and Other Variables." *Journal of Educational Psychology*, 1922, pp. 129-139, 223-235, 277-285.

⁷ R. Byrns and V. A. C. Henmon. "Long Range Predictions of College Achievement." *School and Society*, (1935), 877-880.

⁸ Viola Ames. "Factors Related to High School Achievement." *Journal of Educational Psychology*, (1943), 229-236.

correlation than reported by previous investigations in this field.

The mean correlation between Reading Comprehension and the Intelligence Tests resulted in .78. The mean correlation between the Vocabulary Test and the Otis Intelligence test was .72; the mean " r " of Work-Study Skills and Intelligence was .73; between Language and Intelligence, .60; and the mean " r " between Arithmetic test scores and the Otis scores was .59.

Considering the factor of homogeneity and the small size of the group, these correlations may be considered sizable and significant. Apparently the factors that are operating in the test performance on the *Otis Quick-Scoring Test of Mental Ability*, Beta, are also present to a considerable degree in the Iowa Battery.

Since the Otis Test is largely a reading test, the high relationship between the two tests is to be expected since the Iowa tests are also reading tests. The lowest correlations were obtained between Arithmetic scores and the Otis IQ's. However, even this correlation was as high as most correlations reported in the literature between tests of "Intelligence" and tests of achievement.

The correlations between sub-tests of the same and different forms were computed and compared in Table 4. The relationships between the different forms of Vocabulary and Work Study Skills are higher than the relationships between the different forms of Reading Comprehension, Language, and Arithmetic. The correlations between Reading Comprehension and Vocabulary, and Reading Comprehension and Work-Study Skills are consistently higher than between Reading Comprehension and Language and Reading Comprehension and Arithmetic. Generally the Arithmetic scores correlated less with Arithmetic scores on the other test forms than did scores on the Reading, Vocabulary, Work-Study, and Language Tests.

Sixty-one of the 300 measures computed represent marked relationship; in two instances a correlation of 1.00 was obtained, and in one, a value of $-.22$. The remaining cases represent some degree of relationship.

The Iowa tests show that there is a strong tendency for

TABLE 4
Correlations between Subsequent Test Performances of 27 Arlington, Vermont, Pupils, Grade 6-9-89, on Different Forms of the Iowa Basic Skills Sub-Test Totals, Administered at One-Half-Year Intervals, from June, 1943-June, 1945*

Read. Form N	Read. Form N	76	63	77	76	75	70	79	61	78	51	58	70	65	56	53	58	54	43	62	55	47	46	45	57	52	62
Read. Form O	Read. Form O				76	75	72	72	67	68	62	66	71	73	47	55	62	62	49	27	49	47	46	45	48	57	52
Read. Form L	Read. Form L					65	70	71	69	70	69	66	78	81	56	100	69	69	61	31	42	46	46	45	49	57	52
Read. Form M	Read. Form M					96	80	83	74	59	53	56	65	68	44	46	56	61	41	20	47	49	48	45	41	49	52
Voc. Form N	Voc. Form N						65	75	74	65	55	56	61	62	37	57	69	55	50	49	52	43	43	42	44	49	48
Voc. Form O	Voc. Form O						71	71	74	70	68	66	78	81	56	62	66	61	30	50	58	49	48	45	41	49	48
Voc. Form L	Voc. Form L						67	66	69	69	65	61	77	73	48	69	59	55	54	49	52	43	43	42	44	49	48
Voc. Form M	Voc. Form M						81	84	85	70	56	53	61	63	39	52	64	58	57	44	31	33	33	32	32	33	36
Wkst. Form N	Wkst. Form N						85	73	73	63	62	53	64	68	38	41	61	53	45	16	27	33	34	34	32	33	35
Wkst. Form O	Wkst. Form O							79	79	68	55	54	77	81	40	52	72	69	16	45	23	33	34	34	32	33	35
Wkst. Form L	Wkst. Form L							71	71	61	51	54	64	68	33	35	63	74	16	45	23	33	34	34	32	33	35
Wkst. Form M	Wkst. Form M							67	67	57	49	67	70	73	31	37	76	77	17	45	23	33	34	34	32	33	35
Lang. Form N	Lang. Form N							70	70	60	51	57	86	83	48	51	75	66	17	45	23	33	34	34	32	33	35
Lang. Form O	Lang. Form O							72	72	60	51	54	77	81	40	52	72	69	16	45	23	33	34	34	32	33	35
Lang. Form L	Lang. Form L							71	71	61	51	54	77	81	40	52	72	69	16	45	23	33	34	34	32	33	35
Lang. Form M	Lang. Form M							67	67	57	49	67	70	73	31	37	76	77	17	45	23	33	34	34	32	33	35
Ari. Form N	Ari. Form N							70	70	60	51	57	86	83	48	51	75	66	17	45	23	33	34	34	32	33	35
Ari. Form O	Ari. Form O							72	72	60	51	54	77	81	40	52	72	69	16	45	23	33	34	34	32	33	35
Ari. Form L	Ari. Form L							71	71	61	51	54	77	81	40	52	72	69	16	45	23	33	34	34	32	33	35
Ari. Form M	Ari. Form M							67	67	57	49	67	70	73	31	37	76	77	17	45	23	33	34	34	32	33	35

individuals to maintain relatively the same rank from one half-year to the next. This tendency varies with the subject area in which the pupil is tested. Adequate interpretation of an individual's performance at any given testing period is not possible unless reports of previous testing are available.

Summary and Conclusions

1. This is a report of the continuous testing of 27 Arlington, Vermont, pupils from June, 1943, to June, 1945. The group represents the total grade enrollment of pupils tested at half-year intervals using the various forms of the Iowa Every-Pupil Tests of Basic Skills.

2. The average IQ of the group tested was 104.36 *SD* 11.41.

3. The mean correlations between the Otis results and the sub-test scores were as follows: with Reading Comprehension, .78; with Vocabulary, .72; with Work-Study Skills, .73; with Language, .60; and with Arithmetic, .59. These correlations are considerably higher than those reported in the literature and indicate that the factors which operate in test performance on the Otis are also present in determining the pupil's performance on the achievement tests.

4. Most of the sub-tests reported fairly regular increments from one-half year to the next. However, regressions were noted in Arithmetic and Language tests. An adequate interpretation of the test performance of a group at any period is possible only when the group has been tested previously.

5. Correlations between successive performances on the same sub-test are not high enough to enable an accurate prediction as to rank from one-half year to the next. However, a strong tendency has been noted for individuals to retain their approximate position in the group from one testing period to the next. This tendency varied with subject matter. It was highest with Vocabulary and Work-Study Skills and lowest in Reading, Language, and Arithmetic.

6. Correlations between different sub-tests in most cases revealed a tendency for the individual pupil to maintain the same relative position in the various skills measured. Differences in the strength of this tendency are to be noted according to the subject matter area.

7. Because of the many variables that operate in measurement of human behavior and because of the inaccuracies of the test instruments themselves, an interpretation of the test performances of an individual or group can be cautiously safeguarded only when viewed against preceding performances on the same instrument. Otherwise the hazards of erroneous interpretations are great.

TESTING OCCUPATIONAL TRAINING AND EXPERIENCE

HOMER T ROSENBERGER

U. S. Bureau of Prisons

Introduction

THE testing of occupational training and experience has long presented a difficult problem. All too frequently persons claiming qualifications in an occupation are not able to perform satisfactorily the duties required by that occupation. It is possible for a person to complete specified courses of training with satisfactory marks, to obtain necessary certificates and licenses, and even to obtain satisfactory job ratings, yet be unable actually to perform well the typical duties of the occupation.

In testing occupational training and experience the United States Bureau of Prisons faces many problems which it is attempting to solve through the use of a unique series of "performance tests." The series may prove to be a satisfactory answer to these problems. It is recognized, however, that the Bureau's personnel testing program based on this series has some shortcomings. This testing program needs follow-up and further research before an adequate evaluation of it can be made.

The great variety of activities involved in administering a federal penal institution call for skills associated with at least 35 specialized occupations. The Federal Prison System operates 27 institutions located in 21 states and the District of Columbia. Standards of occupational training and job performance vary from state to state. In the Federal Prison System employees are appointed as Correctional Officers for a probationary period during which they receive the training necessary to give them "prison experience." In the recruit-

ment of Correctional Officers at the entrance salary level the incentive of promotion is offered to those who have qualifying experience for any of these specialized occupational fields when they have acquired experience in prison work.

In a federal prison one finds six "services" or departments, each responsible for a part of the total program of the institution. During a new employee's first week of duty he is given one week of orientation training so that he may become acquainted with the six services and with his fellow workers. During the next three weeks he receives intensive training, based mainly on an assignment to each of the six services. The main objective of this basic training is to teach the trainee (new employee) how to supervise prisoners and how to acquire a working knowledge of the functions of each of the services in the institution.

During the training period new employees are given an opportunity to demonstrate their ability to perform specialized functions. This opportunity is provided by means of performance tests based on 35 occupations. These tests are being administered uniformly in all 27 institutions and were used experimentally in 10 institutions before being used on new employees.

Object of These Tests

The object of the performance tests is to determine by the shortest, simplest and most direct way whether a new employee claiming training or experience in an occupation related to prison work has enough knowledge and skill in the occupation to warrant a two-week trial assignment in the service to which the test applies, to give a reasonable demonstration of his abilities in the occupation. They are devised in such a way that a tradesman can demonstrate, with his own hands and with tools and equipment, proficiency in his trade and a professional person can demonstrate more than mere bookmindedness. The tests, however, make no pretense at evaluating intelligence, educational achievement, judgment or personal adaptability for the occupation to which each test relates. The tests are being used as one step in a training and promotion plan for personnel in the Federal Prison System.

Occupations and Fields of Work to Which the Tests Relate

In the Federal Prison System the related functions performed in an institution are grouped as six "services." The 35 occupations having application to federal prison work and for which performance tests have been prepared are grouped under the six services as follows:

<i>Administrative</i>	<i>Mechanical</i>
Bookkeeper	Auto Mechanic
Clerk Stenographer	Blacksmith
Storekeeper	Bricklayer
Typist	Bus Driver
	Cabinet Maker
<i>Advisory</i>	Carpenter
Librarian	Cement Worker
Recreation Instructor	Electrician
Social Worker	Engineer (Power Plant)
Teacher	Laundryman
	Machinist
<i>Culinary</i>	Painter
Baker	Plasterer
Cook	Plumber
	Sheet Metal Worker
<i>Custodial</i>	Steamfitter
Barber	Shoemaker
Firearms Instructor	Truck Operator
Physical Training Instructor	Welder
<i>Farm</i>	
Dairyman	
Swine Herdsman	
Vegetable Gardener	

Type of Test

Each performance test in the series calls for the actual performance of jobs typical of the occupation to which the test relates. No trick questions or novelties are included. Each test is devised in such a way that a person experienced in the occupation on which the test is based can reasonably be expected to complete the test within a two-hour period.

Typical jobs included in the tests are as follows: for a *Plumber*: caulk a cast iron pipe joint with lead; for a *Firearms Training Instructor*: plan firearms instruction for a one-hour period and actually instruct a group of men after planning the instruction; for a *Bookkeeper*: prepare journal entries to record

specified transactions, post to ledger sheets and prepare a pre-closing trial balance; and for a *Cook*: prepare and serve a meal for four persons.

Jobs to be performed in each test are selected with the view of making them as nearly representative of a wide range of principles involved in the occupation as possible. For instance in a trade such as Painter, mixing of paint, preparing surfaces for painting, painting narrow work, estimating, and staining, each involve an important principle on which the trade is based. Therefore the person taking the performance test for Painter is required to mix paint for particular kinds of surfaces and to match a specified shade, to estimate the number of gallons of paint necessary to put two coats of paint on the exterior of a designated residence, and the like. In order to perform within two hours a wide range of jobs typical of an occupation, the person taking a test such as for Painter is asked to mix and put a priming coat on only two square feet of surface, to prepare but one sash for painting, to putty but one pane of glass, and the like, on the theory that if he can demonstrate familiarity with the operation once he can demonstrate it later in more detail when given a trial assignment on production work.

In preparing each test a careful attempt was made to include items just sufficiently difficult so that a person with inadequate training or low-grade work experience, or one who is "bluffing," cannot perform the jobs satisfactorily. The following typical items taken from the *Performance Test for Carpenter* quickly eliminate men with insufficient training or experience but present no difficulty to a competent Carpenter:

Take a piece of 2" x 6" lumber about 3 feet long and cut one end for the plumb cut (top end) of a one-half pitch roof rafter. On the other end make the seat cut for a rafter of the same pitch to fit over a 4-inch wall plate, and to extend for a 12-inch eave.

Lay a few pieces of finish flooring.

Install one pair of butt hinges (4 half hinges) on 2 pieces of straight 2" x 6" lumber 6 feet 8 inches long, hinging them together as if one were the hanging stile of a door and the other were a door jamb. Give hinges the customary spacing.

With pencil lay out on a board 12 inches wide the cuts necessary for a supporting stringer (stair horse) of a stairway to

rise 9 feet in a run of 12 feet. For the purpose of this item of the test it is sufficient to lay out only two treads and two risers. Determine by your own methods the height of riser and width of tread

In order to be certain that the person taking a performance test has some basic theoretical knowledge of the occupation, a small amount of written work is called for by the test. A trainee taking the *Electrician Performance Test* is asked to draw a wiring diagram for a three-way switch and a wiring diagram for a four-way switch on a lighting circuit. The performance test for Auto Mechanic calls for the writing of the usual reasons why transmissions jump out of gear. The *Steam-fitter Performance Test* calls for a sketch showing fittings and valves in the installation of a steam reducing valve with by pass and an explanation of how water hammer can be avoided in a steam heating system for a two-story dormitory 40' \times 100', connected to a steam line extending to a central steam plant.

How the Tests are Being Used

The performance tests have been developed to measure the qualifications of trainees in the above listed 35 occupations relating to prison work. The tests are a part of the new employee's Basic Training Course. When the employee enters on duty he files information concerning his previous training and work experience. He will be eligible to take one performance test on condition that

- (1) his training or experience or both relate to an occupation on which a test is based.
- (2) he indicates a desire to specialize in that occupation.
- (3) the Bureau considers him qualified to take the test.

When these three conditions are fulfilled the Bureau sends to the institution where the trainee is on duty a copy of the performance test for the occupation concerned, and directs that the test be given the trainee by a person qualified in the occupation represented by the test.

The test is given at the end of four weeks of full-time basic training. If the trainee fails the test he is placed on general custodial assignments. If he passes with a rating of "Good"

or better he is given a two-week trial assignment to special tasks in the occupational field in which he passed the performance test. This two-week trial period is in a sense a continuation of the performance test. During these two weeks the new employee is given an opportunity to demonstrate that he is able to apply his skill and knowledge in the prison environment.

These two weeks constitute an important assignment for observation purposes. During these two weeks of specialized work assignment the trainee is placed on production work in much the same way as a trained employee except that stress is laid on ability to demonstrate a wide range of representative technical skills rather than on supervising prisoners on the more usual operations called for on the particular assignment.

The tasks given the trainee during this two-week trial period are the kind that actually test his ability to do the work himself and that will enable the Supervisor under whom he works to determine the quality of his non-supervisory services. He is assigned to intricate tasks in his field that need to be done, not simply to routine duties. By observing his performance during the test and during this two-week follow-up period of production work the head of the service in which the testing is done can obtain a reasonably good idea of the trainee's abilities and of the likelihood of his qualifying for eventual promotions.

The trainee is rated when he completes the test. The two-week trial period which follows a successful test is simply for the purpose of affording the Supervisor responsible for the particular type of work an opportunity to observe the trainee in actual daily performance. The rating given the trainee as a result of the test is sent to the Bureau of Prisons for record in the Placement Unit, and if his services are satisfactory (rated "Good" or better) during the two week's assignment the Bureau places his name on an eligible list of employees to be given an opportunity to participate in a competitive examination after the necessary prison experience is acquired.

For example, a trainee showing a rather extended training or work experience record in power plant operation and an interest in power plant operation will take a performance test after he has completed his four weeks of Basic Training in

prison work. If he passes the performance test he will be put on technical but non-supervisory production work in the power plant for a two-week period. His test rating and his two-week trial period rating will both be forwarded to the Bureau and he will then be placed on an eligible list for a promotional examination if both these ratings are "Good" or better.

Method of Preparing the Tests

Each of the 35 occupations for which performance tests were prepared were studied in detail—the occupation in civilian life and its counterpart in the Federal Prison System. These data were secured by consulting the United States Employment Service's *Dictionary of Occupational Titles*, and other standard published sources of occupational information, textbooks and manuals giving technical information about trades and professions, by analyses of jobs in the Bureau's institutions, and by consulting specialists on the Bureau's Washington Staff and in its institutions.

In preparing the tests, in order to be fair to *all* persons with appropriate training and work experience, a general practice was made to include only such specific test operations (jobs) as the occupation in civilian life, in all parts of the country, calls for, regardless of whether the counterpart occupation in the Federal Prison System calls for those operations.

As an example: The principal responsibility of a Physical Training Instructor in a Bureau of Prisons institution is to teach Correctional Officers to master defensive tactics, whereas Physical Training Instructors in civilian life are apt to spend much of their time either administering corrective exercises or coaching athletics. Many may have no specific knowledge of defensive tactics. Therefore, in preparing the *Physical Training Instructor Performance Test* no requirement concerning defensive tactics was included. Instead, the test calls for planning physical instruction for a twenty-minute period for various types of persons, and for actually instructing a heterogeneous group for a twenty-minute period. It is assumed that if a trainee can successfully conduct this type of instruction he can readily learn to teach defensive tactics.

In some instances where jobs are typical of an occupation in civilian life but foreign to the counterpart in the Federal Prison System those jobs were not used as test operations. For example, unlike most farmers, employees of the Farm Service in the Bureau's institutions have practically no marketing problem. Their products are consumed in the institutions. Therefore no particular emphasis on marketing operations is found in the three performance tests relating to the Farm Service. In the *Storekeeper Performance Test* no packing and crating or car-loading jobs are included as prison Storekeepers do little shipping. In the *Shoemaker Performance Test* the jobs were confined to the repair of men's work shoes due to the fact that practically all shoe repairing in the Bureau's institutions is confined to men's work shoes.

Along with the selection of jobs that were to be performed in each test a rating system was devised. This was based on the federal government's standard method of rating the efficiency of its employees. This standard rating method was applied to the performance tests due to the fact that it is a method already understood by the supervisory employees who are charged with administering the tests and because of the simplicity of this rating system. Six elements were selected from the standard form used in rating all federal employees. These six elements are used on all 35 tests, with slight exception. The six rating elements are as follows:

Skill in the application of techniques and procedures

Presentability of work (appropriateness of arrangement and appearance of work)

Attention to pertinent detail

Accuracy of final results

Rate of progress on or completion of assignments

Initiative

In order to guide the persons giving the tests, illustrations of both outstanding and weak performance under each of the rating elements were prepared for each of the 35 tests. Persons giving the tests are cautioned that these illustrations are merely *illustrations* and that it is the responsibility of the per-

son giving the test to determine whether the trainee's performance should be rated "outstanding", "weak", or "adequate" on each of the six elements. The following are samples of the illustrations:

AUTO MECHANIC PERFORMANCE TEST

Skill

- outstanding* Used combination socket wrench and screw driver in adjusting valves
weak Used a pipe wrench in tightening and loosening engine head bolt

Pertinent detail

- outstanding* Selected all necessary tools before beginning each job.
weak Lighted matches near oil, grease and gasoline

Initiative

- outstanding* Checked the condition of brake lining and filed down any protruding rivets or recommended that brakes be relined before vehicle is used again.
weak Did not inspect breaker points for pits and burns

Space is provided in the test for jotting down during the test comments concerning the trainee's performance with the thought that such comments will be useful in rating the trainee after the test is completed.

The tests are assembled in two parts. One part, intended for the person giving the test, gives instructions as to how the test shall be given and rated, and provides space for comments on the trainee's performance. The other part is for the trainee who takes the test. It indicates the jobs he is to perform, tells him who will give him the test and provides space for the small amount of written work that is required.

Ten of the tests were tried out experimentally in ten institutions, by correspondence, with the hope that institutional officials would make suggestions for improvement. The results were somewhat disappointing. The next step was to visit several institutions and to try out the tests under carefully controlled conditions. This second attempt to validate the tests was satisfactory and produced two types of improvements, (1) deleting some jobs and adding others to keep the tests within a two-hour completion range and on a completely practical level, and (2) adding illustrations of outstanding or weak performance that were more technically appropriate than some of the original illustrations that were used.

Before each experimental use of any test in the series a careful study of the Bureau's personnel records was made in order to find employees who seemed to possess qualifications in the occupations on which the tests were based. In visiting the institutions this study was supplemented by interviewing all employees considered as candidates for the test. No person was permitted to take more than two tests and in most instances not more than one test. So that persons taking the test would not have an advantage over new employees for whom the tests were being devised, so far as possible the tests were administered only to persons who had never been assigned, since entering the Federal Prison System, to the occupation on which each test was based.

On the basis of experimental use of the tests, by correspondence and "on the ground," revisions were made and the tests were mimeographed. The two-part character of the test and the six-element rating system were retained. The first part of each test, for the person giving the test, averages five pages in length. The part for the trainee, in most tests, is reduced to one page, both sides being used.

Method of Giving the Tests

The tests are given only in the institutions of the Federal Prison System. Each test is administered by the Manager of the service to which it applies or by one of his Assistants. Thus the performance test for Swine Herdsman is administered by the Manager of the Farm Service or by his Assistant in charge of swine, and that for Baker by the Manager of the Culinary Service or by his Assistant in charge of the bakery. Performance tests in any of the shop trades are administered by the Manager of the Mechanical Service or by the Assistant in charge of the shop to which the test relates—sheet metal, welding, and the like. This arrangement assures that each test will be administered only by a supervisory employee qualified in the occupation for which he is testing.

Since it is necessary to observe very closely the performance of each trainee tested, tests are administered to only one trainee at a time. The individual administering the test is completely

responsible for it and for rating the trainee's performance. Performance tests for such occupations as Bookkeeper and Typist may occasionally be administered simultaneously to more than one person, but even here it is probably desirable that these tests be administered to only one person at a time.

Each of the tests in the series requires some preparation on the part of the person who is to administer it. He must first read it through. Then he must secure a suitable place to administer it—a location away from curious eyes. He must assemble whatever equipment and supplies may be needed and see that equipment is in working order so that when the test is started the trainee will not be interrupted. Preparation for most of the tests requires only 15 to 30 minutes but it is essential that it be completed in every detail. The effectiveness of a performance test can be virtually nullified by a combination of two or three neglected details. Failure to start the test in sufficient time to complete it before a shop or other place in which it is given must be closed for the day, and failure to see that there is gasoline in the tank of a mixer that must be operated by the trainee taking the Cement Worker test can cause much annoyance and unsatisfactory test results.

In order to do worthwhile testing it is necessary to put the trainee at ease. He is told that since he is to be tested he must be rated and therefore must be observed, but it is also pointed out to him that we are interested in seeing him do his best on the test. In order to minimize tenseness on the part of the trainee, nothing is said to him or in the test about a time limit. It is also pointed out to him that the test is made up of jobs that are familiar to anyone experienced in the occupation on which it is based and that written work is not overemphasized. This approach usually puts the trainee in the proper frame of mind to get off to a good start on the test, especially if he is not particularly "book-minded."

Unlike paper and pencil tests, this series of tests requires close observation of methods of performance during the testing. It is important that the Supervisor should know exactly how well the trainee handles tools and equipment or how he explains details to a group. Neither his skill, nor his attention to per-

tinient detail nor his accuracy can be determined satisfactorily merely by examining the finished product at the end of the test, even though the final product may be a batch of finished laundry or the test scores attained by a group of men he has instructed.

By jotting down running comments on the trainee's performance, during the test, the person giving the test has an aid at hand when rating. These comments concern things that do not show up in the final product, things that might be forgotten by the time the test is completed, such as special little devices and procedures to attract attention and present a point graphically that may have been used by a trainee in a performance test for an instructor-type occupation.

The person giving the test follows the regular procedure in using the federal government's standard efficiency rating form. He reviews the trainee's performance. He has at hand the illustrations of performance on the six rating elements included in the test and his running comments on the trainee's performance made during the time of the test. In rating the trainee on "*accuracy*" a thorough examination is made of the visible product of the test—such as a paint job, a saw cut, or a gluing job. The amount of time required to complete the test is of course taken into consideration in rating him on "*rate of progress*."

After the person who gave the test has made a study of the trainee's performance in terms of each of the six rating elements he assigns a + (outstanding), √ (adequate), or — (weak) rating on each element. On the basis of these individual ratings he makes a final rating following a formula prescribed on the standard form. The ratings fall into five categories, as follows: Excellent, Very Good, Good, Fair, Unsatisfactory.

Some Preliminary Results

In a typical institution of the 27 in the Federal Prison System, 16 different trainees have taken a total of 18 performance tests, in 12 occupational fields. The ratings were as follows:

Auto Mechanic	—Very Good, Excellent, Excellent, Excellent
Bookkeeper	—Unsatisfactory, Unsatisfactory

Bus Driver	—Very Good
Clerk Stenographer	—Fair
Machinist	—Fair
Painter	—Good
Steamfitter	—Unsatisfactory
Storekeeper	—Good, Very Good
Swine Herdsman	—Good
Truck Operator	—Excellent
Typist	—Fair
Welder	—Good, Excellent

The trainees who received the six ratings of Unsatisfactory and Fair were not eligible for the two-week trial assignment period to prove further their specialized training and experience. Of the 12 trainees who received ratings of Good or better, nine have already been given the two-week trial assignment; eight of them completed it satisfactorily, one did not.

In taking a look at one particular test, the one for Storekeeper, which deals with warehousing, we find the following:

On the basis of training or experience 46 trainees in 14 institutions have qualified, to the satisfaction of the Bureau in Washington, to take the test. The test has been given to 24 of these men in 11 institutions. One of the 24 men received a rating of Excellent, six a rating of Very Good, eight a rating of Good, two a rating of Fair, and the remaining seven failed, receiving a rating of Unsatisfactory. Five of the men who made a rating of Good or better were given the two-week trial assignment and completed the assignment with the following ratings: Good, Good, Good, Very Good, and Excellent, respectively.

Thirteen trainees in 10 institutions qualified for the *Teacher Performance Test*, which has been given to six of these men. They received ratings of Good, Good, Good, Very Good, Very Good, and Very Good, respectively. Three of these six served as full-time teachers in their institutions for a two-week trial, performing during this trial period the duties of one who teaches inmate classes. For the trial period one of the three received a rating of Good, and the other two a rating of Very Good, respectively.

The performance test for Auto Mechanic was given to 10 trainees, a total of 17 in nine institutions having qualified to

take it. One of the 10 who took the test failed it, one received a rating of Good, five a rating of Very Good, and three a rating of Excellent. Six of these nine men who received a rating of Good or better completed their two-week trial period; two were rated Good on the trial period and the other four each, Very Good.

There are several important things to be learned about this testing program. No studies have been made of the correlation between the number of persons failing to qualify to take a performance test and success in general custodial work in the Federal Prison System. No studies have been made of the correlation between failing the tests and resignations from federal prison work. Additional time will be required to determine definitely the reliability of the test items of each test and method of rating.

The flexibility in the method of rating each trainee taking a performance test in the series may at the same time be both a strength and a weakness of this testing program. From a scientific point of view this flexibility is a shortcoming as it precludes testing that is completely objective. This flexibility, resulting from a test score based on a partially controlled personnel rating scale method rather than on a precise weighting scale, makes possible a disparity in rating. The disparity may be reflected when different people with different ideas concerning standards of job performance rate the same trainee on two different tests, in those few instances where a trainee takes a test in each of two occupational fields. It may also be reflected when different trainees in different institutions are being rated on the same test.

Unique Features of the Tests

Each of the 35 tests in the series calls for an actual demonstration and application of specialized knowledge. Whether claiming trade or professional training and experience the person taking a test must demonstrate and apply specialized knowledge basic to the occupational field in which he claims training or experience. In the professional occupations particularly this requirement represents in large measure a novel

approach to testing since access to professional fields is usually gained through the completion of courses in accredited schools or through the passing of state written examinations rather than as the result of a demonstration of ability to meet specific standards of performance.

For the professional occupations included in the group of 35 prison-related occupations it was more difficult to develop this performance test approach than it was for the trades and the other non-professional occupations.

The tests were validated against a broad rating scale based on the quality of performance of jobs that men in the specific occupations are expected to be capable of handling. In measuring an employee's training and experience, or rather the results of such training and experience, such vital matters as the way he goes about performing typical jobs and his speed as related to the regular production rate of a qualified person in the occupation, are taken into consideration.

In this series a newly devised standard test for use throughout an organization is combined with the Federal Government's standard rating system. This makes it possible for supervisory personnel who are not test experts but who are specialists in various occupations to administer the tests, as they are already familiar with both the subject matter of the test and the method of rating. At the same time the use of this type of test makes it possible to give supervisory personnel a large part in the initial selection of employees for their own departments and hence provides an incentive for them to do a thorough testing job.

PREDICTING SUCCESS IN INTRODUCTORY PSYCHOLOGY

LILLIAN G. PORTENIER

University of Wyoming

PREDICTING success in various courses is a perennial problem of all effective advisors of university students. While essential in all fields, such prediction is particularly desirable in Introductory Psychology, since one of the major objectives of the first course in Psychology in most departments is to make it "useful to the individual student by developing his ability to understand and to cope with the personal adjustment problems of daily life" (8).

The present study includes data on 1,035 students enrolled in the introductory course in Psychology at the University of Wyoming for the three years, 1944-1947. During the last two years of this period the testing program for freshmen and other new students at the University was expanded to include a battery of tests. Since 1929 the *Ohio State University Psychological Test* has been administered to all high-school seniors in Wyoming by the Personnel Department of the University and, immediately following their enrollment, to incoming students for whom scores were not on file from the high-school testing program. In addition to the O.S.U. Test, the enlarged battery includes the *American Council of Education Test*, the *Nelson-Denny Reading Test*, the *Iowa High School Content Examination*, the *Wrenn Study Habits Inventory* and the *Cleaton, Kuder and/or Strong Interest Inventory*.

During the school year 1944-1945, in which results on the O.S.U. Test¹ only were available, the correlation between scores on this test and the grades earned in the first quarter's work

¹ Form 21 of the O.S.U. Test was used for all scores throughout the present study.

in Psychology for 173 students was found to be $.667 \pm .026$. This is a higher correlation than has been found in most studies for college grades and various prognostic measures. In 1931 Harl R. Douglass (1) found the computed median for over 250 such correlations on various tests to be .45. L. B. Kinney (3) compiled another summary of prediction studies in 1932. He found a median correlation of .445 between college grades and intelligence test scores. This figure is almost identical with that found by Douglass. Another compilation of studies with similar results was made by David Segel (5) in 1934 working under the U. S. Office of Education. He found a median correlation of .44 between 100 studies on college marks and scores on intelligence tests. Wagner (7) summarized the median correlations between college grades and several specific tests of mental ability. Her correlations for the O.S.U. test range from .43 to .50 and from .17 to .81 with a median between .40 and .50 for the American Council test. Her figures also are very similar to the findings in the earlier studies cited above.

A significantly higher median correlation of .50 or above has been found in summaries of studies made during the past decade. Durflinger (2) states that the higher correlation may be a result of the newer tests of intelligence measuring "more of the factors present in scholastic grades than did the earlier ones, or of college instructors using intelligence test results to assist them in arriving at the grades awarded students." However, intelligence tests vary widely not only among themselves in their ability to predict college grades but any given test may also show wide variations in its ability to predict success in different departments or schools.

Although the correlation of $.667 \pm .026$ between O.S.U. test scores and class marks in the first quarter's work in psychology at Wyoming for 1944-1945 was much higher than that found in most studies, the equation of regression indicated that the O.S.U. score alone is not a reliable basis for prediction of success in individual cases. Consequently, the study was continued the following year using the scores on the new tests in the battery in addition to the O.S.U. scores.

The correlation between grades earned in the first-quarter

course in Introductory Psychology and the various tests and some subtests for 211 students for the year 1945-1946 are shown in Table 1.² The highest correlation, $.672 \pm .025$, was found for grades earned and scores on the O.S.U. Test, and a correlation of $.616 \pm .032$ for grades earned and scores on the *Nelson-Denny Reading Test*. The two tests, Ohio and Nelson-Denny, appear to measure similar abilities as indicated by a correlation of $.762 \pm .020$ between the scores on the two tests (Table 2).

The correlations ($.587 \pm .031$ and $.586 \pm .032$, respectively) between grades earned and scores for each of two subtests of the O.S.U. Test (Analogies and Reading) while almost identical are significantly lower than for the scores on the total test ($.672 \pm .025$).

The scores on the *Nelson-Denny Reading Test* and the *Iowa High School Content Test* also show a high correlation, $.742 \pm .026$ (Table 2), which again seems to indicate a significant overlapping of the abilities measured by the two tests. However, the correlation between grades earned and scores on the Iowa Test, $.487 \pm .045$, is much lower than that found for the Nelson-Denny Test, $.616 \pm .032$. All correlations with the Wrenn Inventory were low.

The American Council Test appears to be definitely inferior to the O.S.U. Test for predicting success in Introductory Psychology as shown by a correlation of $.471 \pm .048$ for scores on the entire test as compared with a correlation of $.672 \pm .025$ for the O.S.U. Test, a difference of .20. The correlation for scores on the subtest for linguistic ability of the A.C.E. Test is slightly higher than for the total test, $.501 \pm .047$ and $.471 \pm .048$, respectively (Table 1). The A.C.E. Test also shows a high correlation with the Iowa and Nelson-Denny tests.

Significant intercorrelations were found between the scores on all four tests, i.e., O.S.U., Iowa, Nelson-Denny, and A.C.E. (Table 2). Since the regression equations indicated that the correlations between the grades and scores on no one of the tests or subtests were sufficiently high for any satisfactory degree of

² The writer is indebted to Miss Vivian Rose Anderson, research assistant, for her aid in compiling these data and for most of the statistical work

TABLE 1

Correlations Between First Quarter Grades in Introductory Psychology, Fall and Winter, 1945-1946, for 211 Students, and Scores on the Various Tests Listed

Test	N	Grades in Psychology
Ohio State University Psychological Test	211	.672 \pm .025
Subtest—Analogies, O.S.U.	211	.587 \pm .031
Subtest—Reading, O.S.U.	211	.586 \pm .032
American Council of Education Test	168	.471 \pm .048
Subtest—Language, A.C.E.	168	.501 \pm .047
Nelson-Denny Reading Test	194	.616 \pm .032
Iowa High School Content Test	178	.487 \pm .045
Wrenn Study Habits Inventory	141	.567 \pm .049

reliability for predicting the success of individual students, multiple correlations were computed as shown in Table 3. The correlation between grades and scores on the O.S.U., Nelson-Denny, and Iowa tests, .682 \pm .029, is no higher than when the scores on only the first two tests are used as variables. The

TABLE 2

Correlations Between Scores on Tests Listed for 211 Students Enrolled for Introductory Psychology, Fall and Winter, 1945-1946

Test	O.S.U.	Iowa H.S. Content	Nelson-Denny
O.S.U.656 \pm .033	.762 \pm .020
Iowa H.S. Content656 \pm .033		.742 \pm .026
Nelson-Denny762 \pm .020	.742 \pm .026	
American Council724 \pm .030	.720 \pm .031	.770 \pm .025

multiple correlations for two and also for three variables are only .01 of a point higher than for the O.S.U. Test alone, (.682 \pm .030 and .672 \pm .025). The regression equations derived from these multiple correlations were also unsatisfactory for individual prediction.

Since the O.S.U. Test gave evidence of being the best single

TABLE 3

Multiple Correlations Between First Quarter Grades in Introductory Psychology, Fall and Winter, 1945-1946, and Scores on the Groups of Tests Listed

Groups of Tests	N	Grades in Psychology
O.S.U., and Nelson-Denny	194	.682 \pm .030
O.S.U., Nelson-Denny, and Iowa High School Content	153	.682 \pm .029

test in the battery for prediction of success, and almost as satisfactory as any two or three tests in the battery, the results were converted into standard scores and analyzed on the basis of the grades earned. The results are compiled in Table 4. Seventy-five per cent of all students with scores on the O.S.U. Test at least two standard deviations above the mean ($+2$ S.D.) earned grades of I in their first quarter's work in Psychology, 16.7 per cent earned grades of II, and the remaining 8.3 per cent, grades of III. For students with O.S.U. scores only $+1.5$ S.D., or above, the percentage of I's earned dropped to 68, and the percentage of II's and III's increased to 20 and

TABLE 4
*Quarter Grades in Relation to Standard Scores on the O.S.U. Test
for 213 Students, Fall and Winter, 1945-1946*

O.S.U. Test Standard Scores	Quarter Grades (Percentage)				
	I	II	III	IV	F
$+2$ S.D. (Above 122)	75.0	16.7	8.3
$+1.5$ S.D. (Above 110)	68.0	20.0	12.0
$+1$ S.D. (Above 98)	48.9	26.7	24.4
$+0.5$ S.D. (Above 87)	35.9	35.9	26.6	1.6	..
-0.5 S.D. (Below 65)	10.6	54.1	21.2	14.1
-1 S.D. (Below 53)	6.5	43.5	28.3	21.3
-1.5 S.D. (Below 41)	36.8	27.4	36.8
-2 S.D. (Below 30)	33.3	50.0	16.7
-1 S.D. to $+1$ S.D. (53 to 93 Inc.)	2.5	31.1	55.7	9.8	0.8
-0.5 S.D. to $+0.5$ S.D. (65 to 87 Inc.)	3.3	32.8	57.4	6.6	..

12 per cent, respectively. While the percentage of I's dropped to 48.9 for students with O.S.U. scores of only $+1$ S.D. or more above the mean, and the percentage of II's and III's increased accordingly (26.7 and 24.4 per cent respectively) none of these students earned grades below III. From these data it would seem safe to predict that students with scores of at least $+1$ S.D., or even $+0.5$ S.D., on the O.S.U. Test should earn grades of III or above in their first quarter's work in Psychology. With O.S.U. scores at least $+1.5$ S.D., two-thirds of the students earned grades of I, and three-fourths earned grades of I if their O.S.U. scores were at least $+2$ S.D.

The above conclusions seem to be verified in general for the grades earned by students enrolled in Introductory Psychology

for the Fall and Winter Quarters, 1946-1947, as shown by Table 5. The enrollment was more than double that of 1945-1946 and the students were taught in classes almost twice as large, i.e., in groups of 75 to 100, rather than 40 to 50. Since a majority of the students were veterans, competition was very keen and the instructors did not have adequate time for individual attention and guidance. Their grading also may have been somewhat more strict as evidenced by a marked reduction in the percentage of I's. No pronounced change was evident in the percentage of III's, IV's and F's. Replacing the basic text

TABLE 5

Quarter Grades in Relation to Standard Scores on the O.S.U. Test for 435 Students, Fall and Winter, 1946-1947

O.S.U. Test Standard Scores	Quarter Grades (Percentage)				
	I	II	III	IV	F
+2 S.D. (Above 122)	50.0	43.8	6.3
+1.5 S.D. (Above 110)	38.3	42.6	19.1
+1 S.D. (Above 98)	29.8	42.9	25.0	2.4	..
+0.5 S.D. (Above 87)	20.3	37.6	37.6	4.5	..
-0.5 S.D. (Below 65)	1.9	15.7	41.0	28.9	12.7
-1 S.D. (Below 53)	0.9	8.8	39.8	32.0	18.4
-1.5 S.D. (Below 41)	10.5	41.7	20.8	27.1
-2 S.D. (Below 30)	14.3	50.0	14.3	21.4
-1 S.D. to +1 S.D. (53 to 98 Inc.)	6.1	25.6	53.3	13.4	1.6
-0.5 S.D. to +0.5 S.D. (65 to 89 Inc.)	9.0	24.6	55.2	9.7	1.5

used in the course from Dockeray, *Psychology*, third edition, used in 1945-1946, to Munn, *Psychology*, in 1946-1947, may have increased the difficulty of the course for some students. However, the curves of distribution for grades earned by the students enrolled for the course were very similar for the two years. Both curves were fairly normal. The percentage of IV's and F's was somewhat less than the percentage of I's and II's resulting in a slight negative skewness of the curves.

Most of the students with negative O.S.U. standard scores earning grades above III, i.e., above average, were veterans who were more mature and far more strongly motivated toward success than the regular underclassmen, a majority of whom were freshmen or sophomores. Case studies were made of all students with negative O.S.U. scores earning grades above

average, and also of all students with positive O.S.U. scores with grades below average. With very few exceptions the discrepancy for the first group of students, i.e., those with negative O.S.U. scores but grades above average appear to be the result of inaccurate O.S.U. scores. The test was administered to most of these students as high-school seniors in the local high schools. In some cases the students were not encouraged to put forth their best efforts and/or inadequate emphasis was placed on the importance of the results as one basis for counseling and guidance. High scores for Psychology on one or more of the interest tests also were indicative of

TABLE 6

First Quarter Grades in Relation to Standard Scores on the A.C.E. Test for 400 Students, Fall and Winter, 1945-1947

American Council of Education Standard Scores	Quarter Grades (Percentage)				
	I	II	III	IV	F
+2 S.D. (Above 148)	33.3	46.7	20.0	. .	.
+1.5 S.D. (Above 137)	29.0	38.7	32.3		.
+1 S.D. (Above 126)	24.6	42.0	31.9	1.5	
+0.5 S.D. (Above 115)	17.0	35.4	41.5	4.8	1.4
-0.5 S.D. (Below 95)	1.8	4.6	47.7	27.5	18.3
-1 S.D. (Below 84)	1.7	41.4	29.3	27.6
-1.5 S.D. (Below 73)	42.9	17.9	39.3
-2 S.D. (Below 62)	27.3	18.2	54.5
-1 S.D. to +1 S.D. (84 to 126 Inc.)	4.8	19.1	55.6	16.7	3.7
-0.5 S.D. to +0.5 S.D. (95 to 115 Inc.)	2.1	18.3	58.5	18.3	2.8

superior achievement for these students. In other cases the discrepancy may be attributed to the greater maturity and experience of the students after their period of service and perhaps also to greater familiarity with tests, which resulted in materially higher scores on the tests taken after enrolling at the University, viz., on the A.C.E., Nelson-Denny, a repetition of the O.S.U., et cetera. Nevertheless, the A.C.E. scores converted into standard scores and analyzed on the basis of the grade earned were found to be less useful for prediction than the O.S.U. scores as shown by the data in Table 6. The findings are not as clear cut and consistent. There is very little difference between the groups with A.C.E. scores of +1 S.D. and above and those with scores of +1.5 S.D. and above in the

percentage of students earning grades of I and II (Table 6). Likewise as many students in the group with A.C.E. scores of -1 S.D. and below were given IV's and F's as in the group with scores of -1.5 S.D. and below. The inferiority of the A.C.E. scores to the O.S.U. scores as a basis for predicting success in Introductory Psychology was indicated also by the lower correlation with grades as shown in Table 1 above, $.471 \pm .048$ vs. $.672 \pm .025$, respectively.

The figures in Table 7 suggest that the O.S.U. scores may be equally useful for predicting success in the second quarter's work in Introductory Psychology. Although the data are

TABLE 7
*Second Quarter Grades in Relation to Standard Scores on the O.S.U. Test
for 176 Students, Winter Quarter, 1946-1947*

O.S.U. Test Standard Scores	Quarter Grades (Percentage)				
	I	II	III	IV	F
+2 S.D. (Above 122)	44.4	33.3	22.2
+1.5 S.D. (Above 110)	20.8	54.2	20.8	4.1	. . .
+1 S.D. (Above 98)	25.5	38.3	29.8	6.4	. . .
+0.5 S.D. (Above 87)	19.2	32.9	42.5	4.1	1.4
-0.5 S.D. (Below 65)	2.0	58.0	28.0	12.0
-1 S.D. (Below 53)	42.3	34.6	23.1
-1.5 S.D. (Below 41)	40.0	20.0	40.0
-2 S.D. (Below 30)	50.0	. . .	50.0
-1 S.D. to +1 S.D. (53 to 98 Inc.) . .	5.8	20.4	57.3	10.7	5.8
-0.5 S.D. to +0.5 S.D. (65 to 87 Inc.) . .	7.7	25.0	46.2	11.5	9.6

based on a similar number of students with a narrower range of O.S.U. scores, the poorer students having been eliminated through failure, the distribution of grades earned is very similar to those for the first quarter's work.

From these data it would seem that the critical O.S.U. score for success in either the first or second quarter's work in Introductory Psychology appears to be about 87 when success is defined as a grade of III or better. The chances for earning grades of I or II, however, are materially increased by raising the critical score to 98 or above. Such critical scores are not only simpler to apply but are also far more practical as a basis for prediction than the regression equations derived from the correlation between grades and test scores suggested earlier.

While the critical scores on the O.S.U. Test in the present study are for Form 21 only, and were determined solely from data on students enrolled in Introductory Psychology as offered at the University of Wyoming during the two-year period 1945-1947, they may be quite generally applicable. A score of 87 is listed as having a percentile rank of 62 on the Ohio College Association norms as compiled by Toops (6) for 3889 college freshmen on Form 21, a score of 98 as a percentile rank of 72. Toops has found a minimum score of 94 for Phi Beta Kappa students at Ohio State University. His studies indicate that most students with scores of 94 or better on Form 21 will do superior university work if properly motivated. This finding has been verified repeatedly in studies (4) made at the University of Wyoming not only for Psychology but also for courses offered in most other departments.

The present study then contributes to the earlier findings primarily in determining more specific critical scores for predicting various levels of success in Introductory Psychology as taught at the University of Wyoming. The scores were derived from data on over one thousand students enrolled in Introductory Psychology over a three-year period. Sections of the course were taught by several different instructors varying widely in their method of presentation. Consequently, the critical scores which appear to have a reasonable degree of reliability and validity for use in counseling students may be quite generally used for predicting success in Introductory Psychology, and they may apply to other fields of study as well.

REFERENCES

1. Douglass, Harl L. *The Relation of High School Preparation and Certain Other Factors to Academic Success at the University of Oregon*. Eugene: University of Oregon Press, 1931.
2. Durlinger, Glenn W. "The Prediction of College Success—A Summary of Recent Findings." *Journal of the American Association of Collegiate Registrars*, XIX (1943), 68-78.
3. Kinney, L. B. *A Summary of the Literature on the Use of Intelligence Tests in Colleges and Universities* (University of Minnesota Committee on Educational Research) Minneapolis: University of Minnesota, 1932.

4. Reed, W. P. and Portenier, L. G. Unpublished data adapted for Wyoming from a table originally published by Herbert Toops, Ohio State University.
5. Segel, David. *Prediction of Success in College*, Bulletin No. 15. Washington: U. S. Office of Education, 1934.
6. Toops, H. A. *Ohio College Association, Bulletin No. 124*. Columbus: Ohio State University, 1940 (mimeographed copy), p. 2410.
7. Wagner, Mazie E. *Prediction of College Performance*. Buffalo: University of Buffalo, 1934, p. 59.
8. Wolfe, Dael. "The First Course in Psychology." *Psychological Bulletin*, XXXIX (1942), 689.

IMPLICATIONS OF A BRIEF STUDY OF PREDICTION OF SUCCESS IN THE MEDICAL SCHOOL, MEDICAL COLLEGE OF VIRGINIA¹

A. W. HURD
Medical College of Virginia

BENJAMIN FINE in his recently published book, *Admission to American Colleges* (4), based on questionnaire returns from 450 American colleges, says that 28 per cent of the colleges use no tests for admission, and that over half make almost no use of admission tests, although 90 per cent say that they use them for guidance and placement. However, many colleges have been using admission tests for several years. The movement is largely an aftermath of World War I. The *Moss Medical Aptitude Test* has been popular in medical schools although there is now a new committee in this field experimenting with other medical aptitude tests. Crawford and Burnham in their new book, *Forecasting College Achievement* (1), point out that there is a need for locally-prepared aptitude tests: "The very fact that numerous . . . universities have constructed their own general intelligence or aptitude tests indicates a recognized need for special, localized evaluations." They add: "To be meaningful, reliability and validity should, whenever possible, be determined locally and anew . . . rather than by mere inference from previously obtained data."

It has been quite conclusively demonstrated that no tests yet devised predict success in college work very reliably. There are, of course, many factors influencing such success and it is, therefore, not surprising that more accurate prediction has not been achieved. In one of our numerous studies at the Medical College, we found that, in one class in the School

¹ Given before the Psychology Section, Virginia Academy of Science, May 9, 1947

of Nursing, a group of students who had dropped out did not differ significantly, in average IQ (*Otis Higher Examination*), or in American Council Psychological percentiles, from those who remained, and we listed 27 possible reasons for the failure of students with inherent ability.

It is becoming clearer that one reason for unreliable prediction resides in invalid and unreliable grades given by instructors. We cannot hope for valid and reliable prediction when the thing we are predicting is, itself, unreliable, invalid, or illy defined. Until we can have reliable and valid measures of achievement, therefore, we cannot have reliable and valid prediction. One of our own studies reported in the *Journal of Educational Research* (5) suggests how this concept works out. When a comprehensive achievement test in Basic Science for Nurses (260 items requiring two hours of work—reliability coefficient $.94 \pm .01$) was used to measure achievement at the end of the preclinical period, the correlation between student ranks on this test and the average ranks in four entrance examinations: *American Council Psychological Examination*, *Otis Test of Mental Ability—Higher Examination*, a local test, *Exercises in Reading Medical Literature*, and a local test in Arithmetic was $.90 \pm .02$. (Spearman rank-difference method used for 29 students in nursing.) This is the highest predictive coefficient I have seen reported in educational literature. It is based on only a few cases but is reported for what it may be worth. A correlation of this magnitude certainly points to the need of further investigation. It gives evidence that with better achievement measures of an objective nature, we may hope for better prediction. Probably much of the unreliability in prediction reported in so many studies is due to poor and uncertain measures of achievement.

This paper presents four correlation coefficients computed for one of our classes in medicine. The coefficient between average grades in three years in the Medical School and percentiles in the *Moss Medical Aptitude Test* for 67 students was found to be $.14 \pm .08$ (*Otis Chart*). That between the grades and a local test in reading medical literature—Comprehension—was $.04 \pm .09$. That between the grades in the Medical

School and the grades received in high school was $.28 \pm .08$. That between the Aptitude Test and the Reading Test was $.47 \pm .08$. There is apparently no reliable prediction from any of the tests used as entrance tests. Even high-school grades are not significant. The two tests—Aptitude and Reading—show a higher relationship. These data taken together suggest a reason for the poor prediction of average grades obtained—abilities recognized in grades given are not the same as abilities represented in the entrance tests. The former undoubtedly are based, partially at least, on social and personality characteristics not represented in the latter. And there are certainly abilities represented in the entrance tests not identified in the grades. Correlations obtained by many investigators are pointing to the conclusion that various of the so-called “new-type” objective tests measure different functions in themselves. True-false, multiple-choice, matching, analogies tests, and the like involve some specific abilities differing from one another. We have some substantiating data for this statement. For example, a true-false test in chemistry gave a correlation coefficient of $.23 \pm .10$ with a controlled completion test in the same area. It seems logical that predicting tests should be of the same sort and test the same abilities as the abilities identified and recognized in the grades given. To predict success in any field—logical reasoning, medicine, surgery, nursing, pharmacy, the pretests for prediction must be closely related in type and abilities sampled to the achievement tests upon which the final grades are based. This implies that achievement ratings which are the mere personal judgments of instructors must be replaced by objective ratings of the sort used as pretests if accurate prediction is to be expected. And pretests should desirably sample the abilities which are identified and recognized in instruction—not merely *general* scholastic or *general* intelligence tests.

Better measures of achievement can only be secured by making sure that we have clear aims and objectives in instruction, with all courses planned to efficiently accomplish these objectives in student performance. So far, our objectives have been too vague and indefinite. Often they are not even

given statement, and then, too often, are expressed for effect in inducing students to attend a particular institution, or to enroll for particular courses. And we cannot overlook the point made by Davis in his article in the *Journal of Educational Psychology* (2): "Although potential abilities rank high among the list of factors indicative of potential success, the individual's active interests constitute the driving force in his achievement."

The movement to identify and to evaluate factors of a comprehensive nature possessed by individuals, like those listed by Holzinger, Thurstone, the Yale Battery, and so on, are undoubtedly valuable for certain purposes but they have not yet arrived at a common denominator and are perhaps too comprehensive and not specific enough for the kind of prediction we seek. As quoted by Crawford and Burnham (1), "On this subject, Godfrey Thomson—offers the following comment: 'The different systems of factors proposed by different schools of *factorists* have each their own advantages and disadvantages, and it is really impossible to decide between them without first deciding why we want to make factorial *analyses* at all.'"

At present, I lean toward the plan of (1) developing institutional and course aims and objectives; (2) pointing instruction toward efficient specific accomplishment of ultimate and immediate objectives; (3) and evaluating preliminary and final student status toward achievement of the abilities represented in the specific purposes in mind. Then we may look forward to purposeful planning, purposeful instruction, and purposeful prediction. As quoted from Tyler in a recent American Council on Education study (6): "Logically, then, the first step in laying the foundation of a broad program of individual evaluation is to formulate a statement of the fundamental purposes for which the school exists."

We have gone forward on this basis in the continuing development of a basic curriculum for nurses (3), and are beginning on a similar basis in the Medical School.

REFERENCES

- 1 Crawford, Albert B. and Burnham, Paul S. *Forecasting College Achievement. A Survey of Aptitude Tests for Higher*

Education. Part I. General Considerations in the Measurement of Academic Promise. New Haven: Yale University Press, 1946.

2. Davis, Robert A. "Testing for Aptitudes." *Journal of Educational Psychology*, XXXVI (1945), 37-45.
- 3 *Educational Research and Nursing Education; What the Testing Program in the School of Nursing has Taught Us; Building a Curriculum for Professional Schools; Building a Course Within a Professional Curriculum.* Richmond: Bureau of Educational Research and Service, 1945-1946
4. Fine, Benjamin. *Admission to American Colleges.* New York: Harper and Brothers, 1946
5. Hurd, Archer Willis. "The Problem of the Prediction of College Success." *Journal of Educational Research*, XXXVIII (1944), 217-219.
6. *Series I. Reports of Committees and Conferences*, No. 20, Vol. VIII. Washington: American Council on Education, 1944.

NEW TESTS*

Army General Classification Test, Civilian Edition, 1947. Three types of test problems are included: vocabulary, arithmetic word problems, and block counting. For high-school, college, and adult groups. Time limit: 40 minutes. Reusable test booklets with answer pad, 45¢, package of 25 answer pads, \$1.65, specimen tests (self-scored), 75¢; test booklets (machine-scored), 35¢, answer sheets (machine-scored) package of 100, \$2.35, scoring keys (machine-scored), 50¢. Published by Science Research Associates.

Arthur Point Scale of Performance Tests, Revised Form II, by Grace Arthur, 1947. A five-part non-verbal test. Standardized on children 5-15 years old, extrapolated norms above 15 years. Knox Cube Test (Arthur Revision), \$2.50; Seguin Form Board (Arthur Revision), \$15.00; Arthur Stencil Design Test I, \$3.00; Porteus Maze Test (Arthur printing), fourteen mazes each in pads of 100 copies, 1-9 pads, 70¢; Healy Picture Completion Test II, \$20.50; Manual, \$2.25; Score sheet pads of 100, \$2.50; package of 100 score cards, \$2.50. Complete set, including all tests, 100 score sheets and manual, \$49.00. Published by The Psychological Corporation

Bennett Use of Library Test, 1947, by Alma Bennett and H. E. Schrammel. This is an achievement test intended to measure knowledge of library organization and practice. Range: High-school and college students. Working time: 50 minutes. Published by Bureau of Educational Measurements.

Cooperative Achievement Tests, 1947, Form X. Test booklets per package of 25 are priced as follows: Group A, \$1.75; Group B, \$2.00; Group C, \$3.00; Group D, \$3.75. Answer sheets per package of 25: special one-sided, 40¢, special two-sided, 60¢; standard one-sided, 40¢; standard two-sided, 50¢. Stencil, 15¢ and 30¢ (for one and two-sided answer sheets, respectively). Hand-scoring key, manual and norms supplied with each order. Published by Cooperative Test Service of the American Council on Education.

* The addresses of the publishers of the tests listed are given at the end of the section. The prices given are for the smallest package listed by the publishers. Lower rates for larger quantities are often available; information on this point may be obtained from the publishers. In some instances, certain details are not included because they were not available at the time of going to press.

American Government Test (Revised Series), by John Haefner. Graphic and verbal material for the purpose of measuring understanding of concepts developed in courses. Working time: 40 minutes. Range: High-school classes. Group A booklets. Special one-sided answer sheets.

American History Test (Revised Series), by Harry D. Berg and H. R. Anderson. Designed for measurement of understanding and information of materials presented in secondary-school or elementary-college courses in American history. Working time: 40 minutes. Range: High-school and elementary-college classes in American history. Group A booklet; special one-sided answer sheets.

Biochemistry Test, by The Cooperative Test Service and the American Chemical Society. Part I planned to measure fundamentals of biological chemistry usually given first semester; Part II, second semester achievement test of a year in Physiological Chemistry. Range: College and medical-school classes in biological chemistry. Working time: 50 minutes for each part. Group B booklet; standard one-sided answer sheets.

Biology Test, by Paul E. Kambly, *et al.* Intended to primarily investigate student's factual knowledge. Working time: 40 minutes. Range: High-school classes in biology. Group B Booklet; special one-sided answer sheets.

Chemistry Test, by Paul J. Burke and Theodore A. Ashford. Planned to test fundamental facts and principles. Range: High-school classes in chemistry. Working time: 40 minutes. Group A booklet; special one-sided answer sheets.

Chemistry Test in Quantitative Analysis, by Cooperative Test Service and the American Chemical Society. Covers "theoretical principles . . . in volumetric or gravimetric procedures in an elementary course . . ." Working time: 110 minutes. Range: college courses in Quantitative Analysis. Group A booklets; standard one-sided answer sheets.

Commercial Arithmetic, by The Staff of the Cooperative Test Service. Norms available for one and two semesters. Working time: 40 minutes. Range: High-school classes in commercial arithmetic. Group A booklet; standard one-sided answer sheet.

Contemporary Affairs for College Students, Form 1947, by E. F. Lindquist, Robert L. Ebel, Goldwin Smith, and Wendell Smith, *et al.* Three parts: Public Affairs, Science and Medicine, Literature and Fine Arts. Working time: 80 minutes. Range: College classes. Group B booklet; special one-sided answer sheets.

Effectiveness of Expression (Higher Level), by Miriam M. Bryan and Geraldine Spaulding. Areas of English investigated by this test include sentence structure and style, diction, and organization. Range Superior 11th- and 12th-grade students and college students. Working time 40 minutes. Group A booklet, special one-sided answer sheets.

Effectiveness of Expression (Lower Level), by Miriam M. Bryan and Geraldine Spaulding. Questions bearing on style, diction and organization of English are included in this test. Range Grades 7 to 12. Working time: 40 minutes. Group A booklet, special one-sided answer sheets.

Foods and Nutrition, by Cooperative Test Service and the Evaluation Committee of the American Home Economics Association. A test which is designed to measure pre-professional information, knowledge of methods and operations in areas relating to home economics and nutrition. Working time. 90 minutes. Range. Pre-professional college classes. Group B booklet, standard two-sided answer sheets.

French Test (Higher Level), by Geraldine Spaulding, Laura Towne, Sarah Wolfson Lorge, *et al.* Consists of three parts, Comprehension, Grammar, Civilization. Range: Students with more than two years high-school French, or more than one year college. Working time: 80 minutes. Group B booklet; special two-sided answer sheet.

French Test (Lower Level), by Geraldine Spaulding, Laura Towne, Sarah Wolfson Lorge, *et al.* Includes sub-tests on Comprehension, Grammar, and Civilization. Range: Students in first two years of high-school or first-year college freshmen. Working time 40 minutes. Group A booklet, special one-sided answer sheet.

General Proficiency Tests, Revised Series:

Mathematics, by Paul J. Burke, Rose E. Lutz, and L. P. Siceloff.

Natural Sciences, by Paul J. Burke, Paul E. Kambly, and V. H. Noll.

Social Studies, by Jeanne M. Bradford, *et al.*

Three separate tests. Each test consists of 2 parts: Terms and Concepts, and Comprehension and Interpretation. Range: Grades 10 through 12 and entering college freshmen. Time: 40 minutes (each test). Booklets are all Group B, special one-sided answer sheets.

General Culture Test (Revised Series), by Norman J. Blair, Jeanne M. Bradford, Miriam M. Bryan, Paul J. Burke, Herbert

Danzei, *et al.* Purpose is to secure an indication of cultural background. Six 30-minute sections comprise the test: Current Social Problems, History and Social Studies, Literature, Science, Fine Arts and Mathematics. Range: College students. Working time: 30 minutes (each section). Group D booklet, answer sheets for 25 tests, \$1.00

General Science Test (Revised Series), by Paul E. Kambly and Paul J. Burke. Information and understanding are understood to be tested. Range: High-school classes in general science. Working time: 40 minutes. Group A booklet, special one-sided answer sheets.

Italian Test (Revised Series), by Peter Riccio, Anthony Cuffari, *et al.* Reading, Vocabulary and Grammar are the three sections of this test. Range: Students with two or more semesters of study of Italian. Working time: 40 minutes. Group A booklet, special one-sided answer sheets.

Mathematics Test for Grades 7, 8, 9, by Vernon Price, *et al.* Four parts labeled Skills; Facts; Terms and Concepts; Applications; and Appreciation are included. Range: Grades 7, 8, 9. Working time: 80 minutes. Group B booklet, special one-sided answer sheets.

Mathematics Pre-Test for College Students, by The Committee On Tests Of The Mathematical Association Of America, *et al.* Intended to offer a sampling of high-school mathematics abilities. Range: Pre-test for beginning college mathematics courses. Working time: 40 minutes. Group A booklet, standard one-sided answer sheets.

Mechanics of Expression, by Geraldine Spaulding and Herbert Danzer. An English test designed to cover usage, punctuation, capitalization and spelling. Range: Grades 7 through 12 and college classes. Working time: 40 minutes. Group A booklet, special one-sided answer sheets.

Modern European History Test (Revised Series), by Frederick H. Stutz, *et al.* Questions call for information and understanding of material from beginning of Renaissance to the present. Range: High-school and elementary-college classes in Modern European History. Working time: 40 minutes. Group A booklet, special one-sided answer sheets.

Physics Test (Revised Series), by Paul J. Burke. Achievement in a secondary-school physics course is expected to be indicated by this test. Range: High-school classes in physics. Working time: 40 minutes. Group A booklet, special one-sided answer sheets.

Science Test for Grades 7, 8, and 9, by Paul Kambly, *et al* Knowledge of principles and functioning of common mechanical devices as well as natural phenomena is called for here. Working time . 80 minutes. Group B booklet, special one-sided answer sheets.

Social Studies Test for Grades 7, 8, and 9, by Harry D. Berg, Elaine Forsyth, *et al*. Content includes items relating to economic, social, physical and historical factors in the development of the American nation. Working time: 80 minutes. Group B booklet, special one-sided answer sheets

World History Test (Revised Series), by Wallace Taylor, Howard R. Anderson, *et al*. "This test covers the entire period from prehistoric times up through the present. . ." Range: High-school classes in World History. Working time 40 minutes. Group A booklet, special one-sided answer sheets.

Diagnostic Reading Tests, by Committee on Diagnostic Reading Tests, 1947. This battery has been developed to ascertain reading skills from Grade VII through first year college. The battery consists of five sections. Two equated forms (Form A and Form B) are available.

The Survey Section, in one booklet, measures (1) the rate of reading general, easy, story-type material, with a check on comprehension; (2) vocabulary comprehension; (3) comprehension of usual textbook-type material.

Section I, Vocabulary, in one booklet, measures vocabulary comprehension. In addition to a general vocabulary score, scores may be obtained on the following types of specialized vocabulary: English grammar and literature, mathematics, science, and social science

Section II, Comprehension, in two booklets: Part 1, *Silent*, measures reading skills used in the usual work-type school assignments, Part 2, *Auditory*, measures ability to comprehend oral presentation of materials

Section III, Rates of Reading: Part 1, General, measures the ability of the student to vary his rate of reading story-type material according to instructions given him; Part 2, Social Science, measures the student's usual rate of reading textbook-type social science material; Part 3, Science, measures the student's usual rate of reading textbook-type science material.

Section IV, Word Attack, in two booklets: Part 1, Oral, includes six graded paragraphs which are read aloud by the student while the examiner records any errors made while the student reads such as mispronunciations, repetitions, substitutions, and others. Part 2, Silent, measures the ability to hear sounds of vowels and consonants and the ability to divide words into syllables. Published by Committee on Diagnostic Reading Tests, Inc. Distributed by Educational Records Bureau.

Differential Aptitude Tests, by George K. Bennett, Harold G. Seashore, and Alexander Wesman, 1947. This battery is intended for the measurement of some isolated factors of ability. Applicable to students in Grades 8 through 12, and young adults.

Verbal Reasoning—Analogies form of item make up this test "aimed at the evaluation of the student's ability to abstract or generalize. . . ." Working time: 30 minutes. Per copy, 13¢; 1-3 packages, \$2.75.*

Numerical Ability—Arithmetic computation is called for in this test. Working time: 30 minutes. Per copy, 10¢; 1-3 packages, \$2.00.*

Abstract Reasoning—Discernment of pattern in serial diagrams and generalizing the operating principle is required here. Working time: 25 minutes. Per copy, 13¢; 1-3 packages, \$2.75.*

Space Relations—Items require "mental manipulation of objects in three dimensional space." Working time: 30 minutes. Per copy, 14¢; 1-3 packages, \$3.00.*

Mechanical Reasoning—Items similar to those of *Bennett Mechanical Comprehension Tests*, i.e., pictured mechanical situations. Working time: 30 minutes. Per copy, 16¢; 1-3 packages, \$3.50.*

Clerical Speed and Accuracy—Intended to measure perceptual speed, brief retention, and response speed. Working time: 6 minutes. Per copy, 13¢; 1-3 packages, \$2.75.*

Language Usage:

Spelling—Adaptable to machine scoring. Working time: 10 minutes.

Sentences—Discrimination of incorrect items of grammar, punctuation and spelling is required. Working time: 25 minutes. Per copy, 13¢; 1-3 packages, \$2.75.*

Answer sheets, package of 50, \$1.50; package of 500, \$13.00; individual report forms, package of 50, \$1.25; stencils, purchased separately, 25¢ per test; one set, \$1.25. Manuals of directions accompany test. Published by The Psychological Corporation.

Farnsworth Dichotomous Test for Color Blindness, by Dean Farnsworth, 1947. An aid in distinguishing the functionally color blind and those moderately defective from normal persons. Materials include rack, color caps, and scoring sheets. Time: About 2 minutes per subject. Set of materials, manual, and 100 analysis sheets, \$21.00; packages of 100 analysis sheets, \$2.50; manual only, 35¢. Published by The Psychological Corporation.

Guilford-Zimmerman Aptitude Survey, by J. P. Guilford and Wayne S. Zimmerman, 1947. This battery is designed to sample seven factorially unique traits of ability in the areas of abstract intelligence, clerical aptitude, and mechanical aptitude. Range: high

* 25 Copies per package

school, college, and adults. Norms are currently available on college groups. Administration time for all seven parts is three hours. Parts of the test may be used in any number or combination. Answer sheets are available for all except Parts III and IV. Manual, 25¢; package of 25, Part I, Verbal Comprehension, Part II, General reasoning, Part III, Numerical Operations, Part IV, Perceptual Speed, \$2.00 each; Part V, Spatial Orientation, and Part VII, Mechanical Knowledge, are \$3.50 per package of 25; Part VI, Spatial Visualization, \$5.00 per package of 25; scoring keys for test booklets, \$1.00 per part. Stencils for answer sheets, 50¢ per part. Published by Sheridan Supply Company.

Harris Tests of Lateral Dominance, by Albert J. Harris, 1947. Hand, eye, and foot dominance are investigated by these tests. Time: 10-15 minutes. Manual including sample copy of Record Form, \$1.50; package of 25 record forms, \$1.50; manual and package of 25 forms, \$2.75. Published by The Psychological Corporation.

Kansas First Year Spanish Test, 1947, by Minnie M. Miller. A 100-item achievement test. Norms are based on high-school students. Working time: 40 minutes. Published by Bureau of Educational Measurements.

Mental Health Analysis, Form A, by Louis P. Thorpe, Willis W. Clark and Ernest W. Tiegs, 1946. Designed to systematically ascertain presence of mental health liabilities and assets. No time limits. Separate tests are available at four levels: Elementary, Grades 4-8; Intermediate, Grades 7-9; Secondary, Grades 9-college; Adult. Package of 25 hand-scoring edition, \$1.75; machine-scoring edition, 7¢ per booklet; answer sheets, 2¢ each; specimen set and accessories (at any level), 35¢. Published by California Test Bureau.

Metropolitan Achievement Tests, New Forms R to V, by Richard D. Allen, Harold H. Bixler, William L. Connor, Frederick B. Graham, and Gertrude H. Hildreth. Measurement of school achievement is the object of these revised batteries, ranging from Grade 1 through the first half of Grade 9. Several forms will be available at each level. Primary I and Primary II Batteries for Grades 1 and 2 respectively, are intended as skill tests. Total time estimated 100 minutes (including several separate testing sessions). Elementary Battery (Grades 3 and 4) contains six subject-matter tests. Total time. 135 minutes. Both Intermediate Battery Complete (Grades 5 and 6) and Advanced Battery Complete (Grades 7, 8, and first half of 9), each contain 10 sub-tests. Time estimates are 200 and 220 minutes, re-

spectively. Partial batteries are available for the Intermediate and the Advanced tests, 6 sub-tests comprising each. Also published separately are arithmetic and reading tests for the Elementary, Intermediate and Advanced Batteries. Packages of 25 tests with manuals, keys, class records and analysis chart for Form R: Primary I, \$1.60; Primary II, \$1.65; Elementary, \$2.25; Intermediate, \$2.70; Advanced, \$2.70; Intermediate Partial or Advanced Partial, \$2.20; Arithmetic Test (elementary, intermediate or advanced), \$1.40 per package; Reading Test, Elementary, \$1.55, Intermediate, \$1.40, and Advanced, \$1.40. Published by World Book Company.

New California Short-Form Test of Mental Maturity, '47 S-Form, by Elizabeth T. Sullivan, Willis W. Clark and Ernest W. Tiegs, 1947. These tests are designed to measure both language and non-language factors of mental capacity. Five separate power tests are available at levels ranging from kindergarten to adult. Tests at the elementary level and above are "self-administering." The tests are administered in one period, with suggested time limits for sub-tests. Forms include Pre-Primary, (Kindergarten and Grade I); Primary (Grades I-III); Elementary (Grades IV-VIII); Intermediate (Grades VII-X and Adult); Advanced (Grades 9-Adult). Package of 25 hand-scoring edition, \$1.20; machine-scoring edition, 25-699, at 6¢ each; answer sheets, 2¢ each; specimen set and accessories (each level), 35¢. Published by California Test Bureau.

Number Fact Check Sheets, Forms A and B, by Roy Cochrane, 1947. Mastery of addition, subtraction, multiplication and division are examined in this test. Entire test printed on front and back of I.B.M. answer sheet. Norms available on Grade V through Grade VIII. Time: about 20 minutes. Hand or machine scoring. Package of 25 answer sheets, 75¢; hand-scoring stencil set, 60¢. Published by California Test Bureau

Ohio Social Acceptance Scale, by The Euclid Elementary Teachers and The College of Education, Ohio State University. Students are required to designate their degree of acceptance of each classmate by choosing the appropriate paragraph description. Issued by Louis E. Rath.

Progressive Tests in Social and Related Sciences, Elementary Battery, Form A, by Georgia Sachs Adams and John A. Sexson, 1947. These subject matter tests are arranged in 3 separate booklets, each composed of 2 sections. Suggested time limits for these power tests are. Part I, Social Studies I, The American Heritage, Peoples of Other Lands and Times, 50 minutes. Part II, Social Studies II, Geography; Basic Social Processes, 60 minutes. Part

III, Related Sciences, Health and Safety; Elementary Science, 40 minutes. Percentile Norms for Grades 4 through 8 Hand-scoring: Parts I and II, \$1.25 per package of 25, Part III, \$1.00 per package of 25. Machine-scoring booklets: Part I and II, 5¢ each; Part III, 4¢ each; answer sheets, 3¢ each. Specimen sets (on any part) and accessories, 35¢ each. Published by California Test Bureau.

Revised Beta Examination, by D. E. Kellogg and N. W. Morton, 1946 Restandardization by R. M. Lindner and M. Gurvitz. This revision makes possible the securing of Beta IQ's similar to Wechsler IQ's. Contents and essentials of administration and scoring remain unchanged. Range: Grade 3 to adult. Time: 30 minutes. Package of 25 with manual and key, \$3.30. Specimen set, 30¢. Published by The Psychological Corporation.

Rich-Engelson Spelling Test, 1947, by Vernita Rich, H. E. Schrammel and Ieleen Engelson. To be used for an indication of spelling accuracy of students in Grades 9 to 12, and college. Time: 30 minutes. Published by Bureau of Educational Measurements.

Science Research Associates Clerical Aptitudes, Form AH, by Richardson, Bellows, Henry, and Company, Inc., 1947. This test is planned to measure three aspects of clerical work: office vocabulary, office arithmetic and office checking. Total time for administration: 35 minutes. Reusable test booklet with answer sheet, 43¢; package of 25 answer pads, \$1.65, package of 25 profile sheets, 45¢; specimen sets, 75¢. Published by Science Research Associates.

Science Research Associates Dictation Skills, by Marion W. Richardson and Ruth A. Pedersen, 1947. For the purpose of testing dictation speed and dictation accuracy; two albums of records are available. No typewriter is required. Accuracy album, \$6.00, speed album, \$9.00; reusable accuracy booklet with answer pad, 45¢, reusable speed booklet with answer pad, 45¢; answer pads for accuracy test, package of 25, \$1.65; answer pads for speed test, package of 25, \$1.65. Published by Science Research Associates.

Science Research Associates Language Skills, Form AH, by Marion W. Richardson and Ruth A. Pedersen, 1947. Designed to measure vocabulary, spelling, word division and punctuation skills. The test is member of a battery of stenographic skills tests, and is expected to operate as a screening instrument in selection. Working time: 20 minutes. Cost: package of 25 test booklets, \$1.90; specimen set, 50¢. Published by Science Research Associates.

Science Research Associates Mechanical Aptitudes, Form AH, by Richardson, Bellows, Henry, and Company, Inc., 1947. Designed to measure separately mechanical knowledge, space relations, and shop arithmetic. Time: 35 minutes. Current norms on high-school graduates attending trade schools. Reusable test booklet with answer pad, 43¢; package of 25 answer pads, \$1.65; package of 25 profile sheets, 45¢; specimen sets, 75¢. Published by Science Research Associates.

Science Research Associates Non-Verbal Form, Form AII, by Robert N. McMurry and Joseph E. King, 1917. Series of pictured objects comprise the items of this test. Scores are dependent on the examinee's ability to reason out differences. No reading is required in the test proper. Time: 10 minutes. Range: 12 to 17 or over. Cost: package of 25 (self-scored) booklets, \$1.90; package of 25 (machine-scored) booklets, \$1.30; package of 100 answer sheets (machine-scored), \$2.35; set of scoring keys (machine-scored), 50¢; specimen sets, 50¢. Published by Science Research Associates.

Science Research Associates Primary Mental Abilities, Preliminary Form, by L. L. Thurstone and Thelma Gwinn Thurstone, 1947. This shortened and simplified version of the Chicago Tests is planned to measure five competent abilities which have been statistically identified as discrete intellectual areas. The five sub-tests—verbal-meaning, space, reasoning, number, and word-fluency—are contained in a single reusable booklet but are separately timed. Administration time estimated at 45 minutes. Range: Ages 11 to 17. Cost: test booklet with answer pad, 43¢; package of 25 answer pads, \$1.65; package of 25 profile sheets, 45¢; specimen set, 75¢. Published by Science Research Associates.

Science Research Associates Reading Record, by Guy T. Buswell, 1947. This test is planned to help determine the present status of the student's reading skill, and to estimate his specific reading strengths and weaknesses relative to educational and occupational goals. Both speed and comprehension are investigated in ten sub-tests published in a single reusable booklet. Administration time: 40 minutes. Applicable to grades 8 to 13. Cost: test booklet with answer pad, 43¢; package of 25 answer pads, \$1.65; package of 25 profile sheets, 45¢; specimen set, 75¢. Published by Science Research Associates.

Science Research Associates Typing Skills, Form A, by Marion W. Richardson and Ruth A. Pedersen, 1947. A part of Science Research Associates Program of Stenographic Skills Tests for hiring or upgrading. Working time: 10 minutes. Cost: package

of 25 test booklets and work sheet, \$1 90; specimen sets, 50¢. Published by Science Research Associates.

Science Research Associates Verbal Classification Form, Form AH, by Thelma Gwinn Thurstone and L. L. Thurstone, 1946. As a short form of the *American Council on Education Psychological Examination* and a revised edition of the *Thurstone Test of Mental Alertness*, this test is designed to measure general intelligence. Range: 12 to 17 or over. Time: 15 minutes. Cost: package of 25 test booklets (self-scored), \$1.90; package of 25 test booklets (machine-scored), \$1.30; package of 100 answer sheets (machine-scored), \$2.35, set of scoring keys (machine-scored), 50¢; specimen set (self-scored), 50¢. Published by Science Research Associates.

Test on Social Usage, by Margaret Stephenson and Ruth Millett, 1947. The purpose is to test fundamentals of courtesy and etiquette. Untimed. Published by McKnight and McKnight.

Thurstone Interest Schedule, by L. L. Thurstone, 1947. A check list yielding 10 scores, Physical Science, Biological Science, Computational, Business, Executive, Persuasive, Linguistic, Humanitarian, Artistic, Musical. Untimed (10 minutes estimated). Applicable at secondary school level and above. Published by The Psychological Corporation.

ADDRESSES OF THE PUBLISHERS AND DISTRIBUTORS OF THE TESTS LISTED

- Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kansas
 California Test Bureau, 5916 Hollywood Boulevard, Los Angeles 28, California.
 Cooperative Test Service of the American Council on Education, 15 Amsterdam Avenue, New York 23, New York.
 Educational Records Bureau, 437 W 59th. St., New York 19, N. Y.
 McKnight and McKnight, 109-111 West Market Street, Bloomington, Indiana.
 Psychological Corporation, 522 Fifth Avenue, New York 18, New York.
 Rath, Dr. Louis E., Director of Research, New York University, School of Education, Washington Square, New York 3, New York.
 Science Research Associates, 228 South Wabash Avenue, Chicago 4, Illinois.
 Sheridan Supply Company, P. O. Box 837, Beverly Hills, California.
 World Book Company, Yonkers, New York.
 World Book Company, Yonkers-on-the-Hudson, New York.

MEASUREMENT ABSTRACTS¹

Aita, J. A., Armitage, S. G., Reitan, R. M. and Rabinovitz, A. "The Use of Certain Psychological Tests in the Evaluation of Brain Injury." *Journal of General Psychology*, XXXVII (1947), 25-44.

This article reports on the diagnostic value of three specific psychological tests in cases of brain injury from military causes: the *Wechsler Mental Ability Scale*, the *Shipley-Hartford Retreat Scale for Intellectual Impairment*, and the *Hunt-Minnesota Test for Organic Brain Damage*. Experimental and control subjects were from Neurological-Neurosurgical Service wards of an Army General Hospital, the former including both focal and diffuse brain damage cases three to six months after injury, and the latter having been classified as cases with no organic cerebral alteration. In general, the authors' findings indicate that the Wechsler M.A.S. seems to have definite value in the determination of brain damage, that the Shipley-Hartford is of doubtful validity, and that most of the Hunt-Minnesota tests show no differentiating ability. *Vernon S. Tracht.*

Anderson, Rose G. "Test Scores and Efficiency Ratings of Machinists." *Journal of Applied Psychology*, XXXI (1947), 377-388.

Supervisors' efficiency ratings and the scores on the following tests were secured for 174 machinists. *Adult Placement Test* (developed by the author), *Bennett Mechanical Comprehension Test*, Form AA, *Revised Minnesota Paper Form Board Test*, and the *MacQuarrie Mechanical Aptitude Test*. The effectiveness of the different tests varied according to the specific-trait. A paper and pencil test of general intelligence, singly or in combination with others, was most discriminating in segregating machinists according to their efficiency ratings. High test scores and low efficiency ratings for twelve of the men indicate the importance of supplementing supervisors' judgments and test scores with personal history and interview data. *Leroy S. Burwen.*

Assum, Arthur L. and Levy, Sidney, J. "A Comparative Study of the Academic Ability and Achievement of Two Groups of College Students." *Journal of Educational Psychology*, XXXVIII (1947), 307-310.

This is a study of the relationship of academic aptitude and achievement in equated groups of adjusted and maladjusted college

students. Aptitude was measured by the *ACE* and *College Reading Ability* and *College Writing Ability* scores, while achievement was measured by the students' grades on comprehensive examinations. The authors detected no significant differences between the groups with respect to aptitude, but critical ratios indicated a significant difference in achievement in favor of the adjusted group. *Harold Mosak.*

Barrett, Dorothy M. and Eaton, Elizabeth B. "Preference for Color or Tint and Some Related Personality Data." *Journal of Personality*, XV (1947), 222-232.

One hundred and fourteen college women were subjects in an experimental investigation of the relationship of color or tint preference to characteristics of personality. Retest reliabilities for 63 subjects for color and tint choices were .83 for tint scores and .79 for color scores. Other tests completed by the subjects included word association, estimation of dots, paper and pencil mazes, and nine attitude, interest, and personality scales. Reliable differences in test results were found for eight of the tests, between subjects preferring colors and those preferring tints. These differences are interpreted as supporting the hypothesis that preference for color is associated with extroversion, and for tint with introversion. *Frances Smith.*

Borow, Henry. "The Measurement of Academic Adjustment." *Journal of the American Association of Collegiate Registrars*, April 1947.

The author describes how the *College Inventory of Academic Adjustment* was developed and validated to meet the need for a measure of non-intellectual aspects in scholastic achievement. Believing that the disparity between the prediction of success (as measured by educational aptitude tests) and demonstrated proficiency (as revealed by earned grades) is in large part attributable to attitudes, aspirations and personal practices outside the scope of such scholastic tests, he devised a 90-item questionnaire compassing the following 6 categories of academic adjustment: curricular, maturity of goals and level of aspiration, personal efficiency, study skills and practices, mental health, and personal relations. All of these sections except the last gave significant, albeit relatively low, correlations with freshman grades. The over-all results indicate that the inventory can be useful in student counseling as an aid in finding sources of academic maladjustment. *Vernon S. Tracht.*

Brozek, Josef, Simonson, Ernst and Keys, Ancel. "A Work Test for Quantitative Study of Visual Performance and Fatigue." *Journal of Applied Psychology*, XXXI (1947), 519-532.

Six subjects were presented with a visual task consisting of the recognition of letters presented in random order on a belt moving behind a narrow slit, the recognized letters being recorded manually

by the subject. Practice periods to attain a performance plateau were followed by a two-hour work period for each subject with three work samples taken during this period. The level of illumination intensity was varied for each of three pairs of subjects. More adequate illumination was found to result in a higher general level of performance, a smaller fatigue decrement in output, and a smaller interindividual variability. The test is recommended as a standard method of studying the effect of different variables on performance and on visual fatigue. *Frances Smith.*

Clarke, Helen J., Rosenzweig, Saul and Fleming, Edith E. "The Reliability of the Rosenzweig Picture-Frustration Study." *Journal of Clinical Psychology*, III (1947), 364-370

The *Rosenzweig Picture Frustration Test* was administered to a group of subjects and then scored by four persons. The points of disagreement between the scorers were analyzed and scoring was revised accordingly. This procedure was carried out three times, increasing the reliability of scoring to a point of approximately 85 per cent agreement between two independent scorers. *Leroy S. Burwen.*

Coleman, William "The Thematic Apperception Test. I. Effect of Recent Experience. II. Some Quantitative Observations." *Journal of Clinical Psychology*, III (1947), 257-264.

The possibility that T.A.T. responses reflect recent experiences was tested by administering to 41 children, ranging in age from 8 to 13½, two approximately equivalent parts of the T.A.T. before and after they had seen a motion picture. Results showed only 1 of the 370 stories, obtained after the film presentation, was apparently directly influenced by the movie's content, suggesting that films of average dramatic intensity do not appreciably affect responses. The author reports his attempts to explore additional problems with the data compiled from the primary investigation, such as the relation of age and sex to mean level of response, characteristics of the most and the least productive pictures, predominant emotional tone of responses for the various pictures, and finally, changes in emotional tone from plot to story ending *Vernon S. Tracht.*

Davis, Robert A. and Rood, Edward J. "Remembering and Forgetting Arithmetical Abilities." *Journal of Educational Psychology*, XXXVIII (1947), 216-222.

The authors, by administering the *Schorling-Clark-Potter Arithmetic Tests* to 56 junior high-school pupils, investigated whether basic arithmetic skills are preserved while more complex material is being studied. Initial administration of the test occurred upon entrance into the seventh grade, and four times subsequently while the pupils were learning more complex material such as operations involving the use of percentages. Pupils evidenced progressive improvement upon each retest, but not uniformly for all operations, subtraction

and multiplication showing less improvement than division, fractions, and problems. *Harold Mosak.*

Dvorak, Beatrice J. "The New USES General Aptitude Test Battery." *Journal of Applied Psychology*, XXXI (1947), 372-376.

The author describes the manner in which the new *USES General Aptitude Test* battery was developed and standardized. He also describes the types of occupational aptitude patterns indicated by the test, and the general uses of the battery. Also pointed out are some of the limitations of the test and the need for further development. *Leroy S. Burwen.*

Gray, Horace. "Psychological Types and Changes with Age." *Journal of Clinical Psychology*, III (1947), 273-277.

Jung's psychological types were examined by means of a questionnaire administered at varying times to 1000 subjects ranging in age from 10 to 80 and equally divided as to sex. The specific aim of the study was to ascertain differences in type revealed at different ages. Calculating proportional frequency of extraverts, intuitives, and feeling-type individuals in each five-year age group, and averaging of ratings within each age group, both support the conclusion that there is a slight progressive decrease with age in the feeling-function, definite decrease in extraversion, and most marked decrease in the intuition-function. *Frances Smith.*

Hall, William E. and Cushing, James R. "The Relative Value of Three Methods of Presenting Learning Material." *Journal of Psychology*, XXIV (1947), 57-62.

This study first suggested itself during the recent war when the authors attempted to justify the Army's expenditures for movie tone film for instructional purposes. In an effort to measure the relative effect of film, lecture, and reading, the authors selected three films, one informational, one theoretical, and one demonstrational, and prepared them in lecture and mimeographed form. Administering these to three groups in controlled order, the authors found that no one method was significantly superior and suggest that "superiority in learning results from the relationship of the peculiar qualities of the learner to the method and is not inherent in the method itself." *Harold Mosak.*

Havighurst, Robert J. and Breese, Fay H. "Relation between Ability and Social Status in a Midwestern Community: III. Primary Mental Abilities." *Journal of Educational Psychology*, XXXVIII (1947), 241-247.

A comparison of scores on the *Primary Mental Abilities Test* made by all children born in 1932 in the three lowest social groups in a mid-western community is "consistent with the hypothesis that

high social position corresponds to high ability," the lowest social group scoring definitely low on all tested abilities. This relationship is most pronounced between social status and the Number, Verbal, and Word Fluency factors. Sex differences were generally in favor of girls except for the Space factor where boys were superior, and the Verbal factor where no significant differences were discovered. *Harold Mosak.*

Kriegman, George and Hansen, Frank W. "VIBS: A Short Form of the Wechsler-Bellevue Intelligence Scale." *Journal of Clinical Psychology*, III (1947), 209-216.

To meet the need for a brief, easily administered, quantitative estimate of mental ability, 4 sub-tests of the *Wechsler-Bellevue Intelligence Scale*—Vocabulary, Information, Block Design and Similarities—were chosen to form a new test. The subjects were soldiers, approximately evenly distributed as to race and to urban and rural environments, of whom 409 were given the 4 sub-tests (VIBS) and 89 the full Wechsler Scale. In addition the Clinical records of 118 patients, referred for emotional problems, were used; none, however, was psychotic in any of the groups in the study. The VIBS was found to be a quick, reliable and valid measure, especially helpful in differentiating normals from suspected mental defectives. *Vernon S. Tracht.*

Lough, Orpha M. "Women Students in Liberal Arts, Nursing, and Teacher Training Curricula and the Minnesota Multiphasic Personality Inventory." *Journal of Applied Psychology*, XXXI (1947), 437-445.

The MMPI was administered to 300 unmarried college women divided into four groups: 54 liberal arts students, 61 cadet nurses, 74 elementary education majors, and 111 music education majors. All groups were comparable in terms of age. Results show all groups to be relatively stable with a slight tendency toward hypomania. There were no significant differences between groups on the separate scales. Nurse cadets in general had more masculine interests and were less emotional than the other groups. The author concludes on the basis of this and other studies that the MMPI has little or no value in educational selection. *Leroy S. Burwen.*

Luchins, Abraham S. "Proposed Methods of Studying Degrees of Rigidity in Behavior." *Journal of Personality*, XV (1947), 242-246.

Techniques employed in experimental study of rigidity of behavior are described under the following headings: arithmetical problems, hidden-word tasks, pencil-and-paper mazes, series of drawings in which one design evolves through ambiguous stages into another, tapping rhythm, and set for color. Data obtained in various studies with normal children and adults and in a study by the author of

soldiers in neuropsychiatric hospitals, indicate that the hospital patients showed more rigidity. It is suggested that further systematic experimentation with the described and similar techniques will be of value both for clinical work and for the theory of personality dynamics. *Frances Smith.*

Malter, Morton. "The Ability of Children to Read Cross-Sections." *Journal of Educational Psychology*, XXXVIII (1947), 157-166.

In one of a series of studies on audio-visual aids, the author required three groups of elementary-school pupils to match eight items and cross-sections of these items. A comparison of observed and theoretical frequencies of scores demonstrated that the observed scores could not be accounted for solely on the basis of chance, and necessitated the rejection of the hypothesis that pupils are unable to read cross-sections. An important implication of this study is that children can be taught to read cross-sections as early as the fourth grade. *Harold Mosak.*

Malter, Morton. "The Ability of Children to Read a Process Diagram." *Journal of Educational Psychology*, XXXVIII (1947), 290-298.

Employing a diagram illustrating the flow of wheat in a flour mill, the author investigated whether elementary-school pupils were capable of following the operations in diagrams of industrial processes. Results obtained from 227 children from two schools indicated that these pupils were not able to follow the flow of materials in the diagram. However, 34 pupils from a third school were aided when the diagram was altered to include "signposts" indicating the beginning and terminal points and directions to follow the arrows. *Harold Mosak.*

Morris, Woodrow W. "A Preliminary Evaluation of the Minnesota Multiphasic Personality Inventory." *Journal of Clinical Psychology*, III (1947), 370-374.

The MMPI scores of 320 naval personnel referred to the Neuropsychiatric Service were studied in a preliminary evaluation of the test as a diagnostic instrument. Average scores of patients classified as "No Disease" and of patients in three pathological categories were all within the normal limits on Mf, Pa, and Ma scales, while all abnormal patients scored significantly above 70 on D. The Inventory was found to differentiate between borderline normals and serious pathological states, but not between pathological groups. It is recommended that the Inventory be used with caution as a diagnostic instrument until a further critical evaluation is completed. *Frances Smith.*

Rabin, A. I. and Sanderson, M. H. "An Experimental Inquiry into Some Rorschach Procedures." *Journal of Clinical Psychology*, III (1947), 216-225.

Of a group of 34 student nurses, 17 (group I) were given the Rorschach in standard order and 17 (group II) were given it with the order of the cards reversed. Two months later the procedure was repeated, except that this time group I was given the reversed order and group II was given the standard order. From these data the authors present conclusions relating to the order of cards, "experience balance," difficulty of cards and color shock. *Leroy S. Burwen.*

Remmers, H. H. and Weltman, Naomi. "Attitude Inter-Relationships of Youth, Their Parents, and Their Teachers" *Journal of Social Psychology*, XXVI (1947), 61-68.

To test the broad hypothesis so prevalent in our Western culture that youth is inevitably in conflict with his elders, and specifically to find how greatly youths' attitudes vary from those of their parents and teachers, the Purdue Opinion Poll for Young People No. 7 was given to 207 students from 10 senior high schools in rural Illinois and Indiana and to their parents and teachers. In addition to the 16 opinion items, the poll contained questions on home environment, political preference, school grade, sex, and place of residence. Results indicate a high degree of similarity of opinion between parents and children, less so between teachers and their students or between teachers and parents. *Vernon S. Tracht.*

Sherman, Arthur Wesley, Jr. "Emancipation Status of College Students." *Journal of Genetic Psychology*, LXVIII (1946), 171-180.

This describes the development of a questionnaire, suitable for older adolescent college students of both sexes, which would indicate whether a subject was emotionally free from his parents. Sixty items were devised and scored with the idea in mind of securing a single numerical index of emancipation status. Subjects qualifying for the study included 238 men and 200 women, all from classes in psychology and sociology, the majority being under 20 years of age. The relation of emancipation status to other variables (such as sex, age, length of residence in college and socio-economic status) was studied by a comparison of the most and least emancipated groups, the greatest difference between them being found in sex. Since the other variables showed meager relations, the author suggests that future studies concern themselves with the parent-child relationship and with parental controls exercised therein. *Vernon S. Tracht.*

Triggs, Frances O. "Critique of Van Allyn's System of Vocational Counseling." *Journal of Applied Psychology*, XXXI (1947), 536-544.

The author states that the *Job Qualifications Inventory* and the

Job Placement Reference as developed and used by Van Allyn in his system of vocational counseling need further investigation. The JQI is concerned primarily with sustained, verified interest, which Van Allyn feels is the main criterion for good vocational counseling, although he points out that minimal intelligence and physical standards for any job must be met. Triggs concludes that the theory and the instrument implementing it both require further validation. *Leroy S. Burwen.*

Triggs, Frances O. "The Measured Interests of Nurses." *Journal of Educational Research*, XII (1947), 25-34.

To supplement ability and aptitude ratings in vocational counseling, the author employed the *Kuder Preference Record* to measure the interests of women in the nursing profession. The Preference Record, administered to 1429 women from various occupational groups and 826 nurses, demonstrated significant differences at the 1 per cent level in all interest areas except the Artistic where the difference was at the 5 per cent level of significance. Differences were positive (i.e., in the direction of the nurses) on the Social Service, Scientific, Artistic, and Musical scales and negative on the Clerical, Computational, and Literary scales. *Harold Mosak.*

Witty, Paul. "Reading Retardation in the Secondary School." *Journal of Experimental Education*, XV (1947), 314-317.

Studies demonstrate that while formal instruction in silent reading is discontinued in the fifth or sixth grade, many pupils have not developed fundamental reading skills and are as a result handicapped in the acquisition of knowledge in the history, science, and language areas. Intelligence and reading tests were administered to one hundred ninth-grade pupils referred as reading problems to the Northwestern University Psycho-Educational Clinic. Analysis of the data reveals the following factors—poor reading habits, physical defects, deficient interest, and emotional maladjustment—as characteristic of this group. Suggestions for the establishment of remedial reading classes in the secondary school are included. *Harold Mosak.*

ADDITIONAL ARTICLES NOT ABSTRACTED

Allen, Robert M. "The Test Performance of the Brain Injured." *Journal of Clinical Psychology*, III (1947), 225-230.

Arkola, Audrey L. "An Experimental Study of the Effects of Sodium Amytal Upon Performance of the Hunt-Minnesota Test for Organic Brain Damage." *Journal of Clinical Psychology*, III (1947), 392-396.

Asher, E. J. and Kahn, David. "The Effect of 'Look' and 'Read' Directions Upon the Attention Value of Illustrations and Texts in Magazine Advertisements." *Journal of Applied Psychology*, XXXI (1947), 431-436.

- Berg, Irwin A. "A Study of Success and Failure Among Student Nurses." *Journal of Applied Psychology*, XXXI (1947), 389-396.
- Blanchard, B. Everard. "A Social Acceptance Study of Transported and Non-Transported Pupils in a Rural Secondary School." *Journal of Experimental Education*, XV (1947), 291-303.
- Brower, Daniel. "The Relations Between Minnesota Multiphasic Personality Inventory Scores and Cardiovascular Measures Before and After Experimentally Induced Visuo-Motor Conflict." *Journal of Social Psychology*, XXVI (1947), 55-60.
- Brown, C. W. and Ghiselli, E. E. "Factors Related to the Proficiency of Motor Coach Operators." *Journal of Applied Psychology*, XXXI (1947), 477-479.
- Cruickshank, William M. "Qualitative Analysis of Intelligence Test Responses." *Journal of Clinical Psychology*, III (1947), 381-386.
- Davis, Clifford E. "The Minnesota Multiphasic Personality Inventory: A New Method of Scoring and Analysis." *Journal of Clinical Psychology*, III (1947), 298-301.
- Deabler, Herdis L. "The Psychotherapeutic Use of the Thematic Apperception Test." *Journal of Clinical Psychology*, III (1947), 246-252.
- Edgerton, Harold A., Britt, Stuart H. and Lemmon, William B. "Reliability of Anecdotal Material in the First Annual Science Talent Search." *Journal of Applied Psychology*, XXXI (1947), 413-424.
- Foster, Austin. "Age and the Wechsler-Bellevue Scattergraph." *Journal of Clinical Psychology*, III (1947), 396-397.
- Garfield, Sol L. "The Rorschach Test in Clinical Diagnosis." *Journal of Clinical Psychology*, III (1947), 375-381.
- Goodman, C. H. "The MacQuarrie Test for Mechanical Ability: III, Follow-up Study." *Journal of Applied Psychology*, XXXI (1947), 502-510.
- Hampton, Peter J. "The Minnesota Multiphasic Personality Inventory as a Psychometric Tool for Diagnosing Personality Disorders Among College Students." *Journal of Social Psychology*, XXVI (1947), 99-108.
- Horrocks, John E. and Kinzer, John R. "The Construction of a Form-Board to Measure Spatial Relations—Dexterity." *Journal of Psychology*, XXIV (1947), 89-91.
- Jackson, Joseph. "A Note on the Crystallization of Vocational Interests." *Journal of Social Psychology*, XXVI (1947), 125-130.
- Krise, E. Morley. "A Short Method of Scoring the Minnesota Multiphasic Personality Inventory." *Journal of Clinical Psychology*, III (1947), 386-392.
- Lawshe, C. H., Jr. and Wilson, R. F. "Studies in Job Evaluation. 6. The Reliability of Two Point Rating Systems." *Journal of Applied Psychology*, XXXI (1947), 355-365.

- Link, Henry C. "What Does Americanism Mean to the American People?" *Journal of Applied Psychology*, XXXI (1947), 425-430.
- MacCrone, I. D. "Reaction to Domination in a Colour-Caste Society: A Preliminary Study of the Race Attitudes of a Dominated Group." *Journal of Social Psychology*, XXVI (1947), 69-98.
- Norman, Ralph D. "A Comparison of Earlier and Later Success in Naval Aviation Training." *Journal of Applied Psychology*, XXXI (1947), 511-518.
- Owens, William A. "Item Form and 'False-Positive' Responses on a Neurotic Inventory." *Journal of Clinical Psychology*, III (1947), 264-269.
- Preston, Malcolm G. "Trait Variability as a Function of Practice and of Age." *Journal of General Psychology*, XXXVII (1947), 3-14.
- Reiner, William B. "The Value of Cause and Effect Analysis in Developing Ability to Recognize Cause and Effect Relationships." *Journal of Experimental Education*, XV (1947), 324-330.
- Rothe, H. F. "Distribution of Test Scores of Industrial Employees and Applicants." *Journal of Applied Psychology*, XXXI (1947), 480-483.
- Rothe, H. F. "Output Rates Among Machine Operators: I. Distributions and Their Reliability." *Journal of Applied Psychology*, XXXI (1947), 484-489.
- Sisk, Henry L. "A Reply to Winfield's Study of the Multiple Choice Rorschach." *Journal of Applied Psychology*, XXXI (1947), 446-448.
- Snidecor, John C. and Hanley, Theodore D. "The Construction of a Test of Ability to Repeat Spoken Messages." *Journal of Applied Psychology*, XXXI (1947), 397-405.
- Tilton, J. W. "An Experimental Effort to Change the Achievement Test Profile." *Journal of Experimental Education*, XV (1947), 318-323.

THE CONTRIBUTORS

Leo Chalfen—M.S., University of Pittsburgh, 1945; candidate for Ph.D., New York University. Intern in Clinical Psychology, Western State Psychiatric Institute and Clinic, 1943–1945. Psychologist and Vocational Counselor, Philadelphia Employment and Vocational Bureau, 1945–1947. Group Guidance Counselor, Group Guidance Service, Philadelphia, 1946. Associate, American Psychological Association. Member, American College Personnel Association, National Vocational Guidance Association.

Benjamin Fruchter—M.A., University of Southern California, 1946. With the AAF Aviation Psychology Program, 1942–1945. Vocational Counselor, University of Southern California, Veterans' Guidance Center, 1945–. Teaching Assistant in Psychology and General Studies, University of Southern California, 1946–. Contributor to the *Volume on Printed Tests* of the Aviation Psychology Series (soon to be published). Associate Member, American Psychological Association, Psychometric Society.

Archer Willis Hurd—Ph.D., University of Minnesota, 1928. Instructor, University of Chicago, summers 1925–1926. Assistant Director, Bureau of Educational Research, University of Minnesota, 1926–1929. Research Associate, Teachers' College, Columbia University, 1929–1934. Associate Professor of Education, Northern Montana College, 1935–1936. Professor of Education and Dean of the College, Hamline University, 1936–1943. Director, Bureau of Educational Research and Service, Medical College of Virginia, 1943–. Visiting Professor of Education, West Virginia University, 1936, 1937. Visiting Professor of Education, University of Missouri, 1935, 1939. Visiting Professor of Education, University of Illinois, summer 1940. Author of numerous books and articles on testing, educational psychology, and research. Member, American Association for the Advancement of Science (Fellow), American Association of University Professors, American Educational Research Association, National Association for Research in Science Teaching (President, 1935), National Education Association, American Association of School Administrators, National Association of College Teachers of Education, National Society for the Study of Education, Virginia Academy of Science, Society for the Advancement of Education.

George S. Klein—Ph.D., Columbia University, 1942. Assistant, Columbia University, 1939–1941. Fellow and Teaching Fellow, Col-

lege of the City of New York, 1939-1942. Instructor, Brooklyn College, 1941-1942. With the U. S. Army: Psychological Assistant, Psychological Unit No. 1, AAF Aviation Psychology Program, 1942-1944; Aviation Psychologist, Psychological Research Project (Radar), AAF Aviation Program, 1944-1946. Staff Psychologist, The Menninger Clinic; Assistant Professor, University of Kansas and Menninger Foundation School of Clinical Psychology, 1946-. Author of articles on perception, personality, and clinical tests in the *Journal of Abnormal and Social Psychology*, *Archives of Psychology*, *Journal of General Psychology* and in Army publications.

W. C. Kvaraceus—Ph.D., Harvard University. Assistant Superintendent of Schools and Director of Guidance, Research and Curriculum, Passaic, N. J., Public Schools. Director of Guidance, Research and Curriculum, Brockton, Mass., Public Schools. Personnel Technician in Classification and Replacement Branch, AGO; Industrial Personnel Division, ASF, during the war. Author of *Juvenile Delinquency and the School*.

Mary A. Lanigan—M.A., Boston University. Public school teacher in Ludlow, Vermont; Bennington, Vermont; and in the Arlington High School, Arlington, Vermont. Graduate Assistant, Boston University. Assistant to the Superintendent of Schools in Bennington, Vermont, at present.

Glenn C. Martin—M.S., University of Washington, 1935. Graduate student, University of California, 1935-1936. With the U.S. Army Air Forces, Psychological Research, 1942-1945. Principal Psychologist, Veterans' Counseling Service, Los Angeles Board of Education, 1945-1947. Instructor in Psychology and Philosophy, Santa Monica College, 1947-. Member, American Psychological Association, Western Psychological Association, National Education Association.

Lillian G. Portenier—Ph.D., Columbia University, 1933. Professor of Psychology, College of Liberal Arts, University of Wyoming, 1930-. Acting Head, Department of Psychology, and Acting Director of Personnel, University of Wyoming, 1942-1945. Vice-President, International Council of Women Psychologists, 1947-1949. Executive Secretary, Rocky Mountain Branch, American Psychological Association, 1944-1948. Fellow, American Psychological Association. Member, American Association of University Professors.

Homer T. Rosenberger—Ph.D., Cornell University, 1932. Teacher in high school, Tidioute, Pa., 1930-1931. Teacher at Susquehanna University, summer 1933. Teacher in Adult Night School, Lock Haven, Pa., 1933-1935. Educational research and administration, U. S. Office of Education, 1935-1942. Member, Civilian Conservation Safety Council, representing U. S. Office of Education,

1938-1942. Supervisor of Training, U. S. Bureau of Prisons, 1942-. Author of articles in *School Life* and *The Social Studies for Teachers and Administrators*. Member, Phi Delta Kappa, Pi Gamma Mu, Training Officers Conference (Steering Committee), Pennsylvania Historical Association (Council), Pennsylvania Prison Society (Acting Committee).

Verner M. Sims—Ph.D., Yale University, 1926. High-school teacher, Kentucky and Connecticut, 1920-1926. Professor of Education and Psychology, Louisiana Polytechnic Institute, 1926-1928. Associate Professor of Psychology, University of Alabama, College of Education, 1928-1939. Professor and Staff member, Bureau of Educational Research, University of Alabama, 1939-. Member, American Psychological Association, American Educational Research Association.

Stanley A. Stromswold—M.A., University of Minnesota, 1946. Counselor and Assistant to the Dean of Students' Work, Arts College, University of Minnesota, 1946-1947. Instructor and Administrative Officer in charge of the Junior College Counseling Office, University of Minnesota, 1947-1948. Member, American Society for the Study of Education, Phi Delta Kappa, American College Personnel Association.

Arthur E. Traxler—Ph D., University of Chicago, 1932. High School Principal and Superintendent in Kansas schools, 1920-1928. Psychologist, University of Chicago High School, 1931-1936. Research Associate, Educational Records Bureau, 1936-1938, Assistant Director, 1938-1941, Associate Director, 1941-. Summer and part-time teaching, University of Chicago, University of Arkansas, University of Alabama, Columbia University, Temple University, University of California. Author of reading tests and textbooks for use in the teaching of reading in junior and senior high schools. Series of publications on measurement and guidance issued by the Educational Records Bureau. Author of a textbook on guidance. Contributor of articles to various educational and psychological journals. Member, American Association for the Advancement of Science, American Educational Research Association, Phi Delta Kappa, Kappa Delta Phi, Psychometric Society. Associate, American Psychological Association.

C. Gilbert Wrenn—Ph.D., Stanford University, 1932. Vocational Counselor, Stanford University, 1928-1936. Associate Director, General College; Associate Professor of Educational Psychology, 1936-1938; Professor of Educational Psychology, 1938-, University of Minnesota. Consultant, Student Personnel Teacher Education Commission of the American Council on Education, 1939-1942. On military leave with the U. S. Armed Forces, Personnel Officer in Bureau of Naval Personnel and Pacific Area, 1942-1946. Associate,

American Youth Commission, 1939-1941. Author and co-author of *Student Personnel Problems*, *Studying Effectively*, *Aids to Group Guidance*, *Time on Their Hands*, and numerous journal articles. President, National Vocational Guidance Association. Vice-President, Council of Guidance and Personnel Association, 1946-.

EDUCATIONAL and
PSYCHOLOGICAL



MEASUREMENT

VOLUME EIGHT, NUMBER TWO, SUMMER, 1948

<i>Are Vocational Orientation Courses Worth Their Salt?</i> C. HAROLD STONE.....	161
<i>An Optometric Aptitude Test.</i> NEIL D. WARREN and A. A. CANFIELD.....	183
<i>Selection and Evaluation of West Point Cadets.</i> DONALD E. BAIER..	193
<i>On the Role of Directive and Non-Directive Techniques in the Counseling Process.</i> JOHN M. BUTLER ..	201
<i>Counseling: Eclectic or Systematic?</i> RAY H. BIXLER.....	211
<i>An Application of Guttman's New Scaling Techniques to an Attitude Questionnaire.</i> KENNETH E. CLARK and PHILIP H. KRIEDT	215
<i>Comparisons of Self-Rating, Peer-Ratings, and Expert's-Ratings of Personality Adjustment.</i> MARGARET G. POWELL...	225
<i>A Review of Some Problems Related to the Mail Questionnaire Technique.</i> RALPH D. NORMAN....	235
<i>Personality Factors in the Psychological Weaning of College Women.</i> ARTHUR W. SHERMAN, JR. .	249
<i>Note Concerning "The Validity of Standard and Custom-Built Personality Inventories in a Pilot Selection Program," by Donald E. Super.</i> DONCASTER G. HUMM..	257
<i>Measurement Abstracts</i>	263
<i>The Contributors.</i>	275

ARE VOCATIONAL ORIENTATION COURSES WORTH THEIR SALT?

C. HAROLD STONE

University of Minnesota

ARE courses in occupations a substitute for individual counseling? Can we achieve desired ends in a student personnel program by lightening the counselor's load through referral to him of only the problem cases, and by "counseling" the majority of students through the medium of group guidance classes? What, if anything, do vocational orientation courses contribute to the facilitation of the individual counseling program? Which is better—a well-formulated sequence of courses in vocational orientation or a comprehensive program for counseling students individually? Or, are there worthwhile values to be gained through an integration of individual counseling and group guidance functions? Definitive answers to these questions would carry us a long way toward a clearer understanding of what procedures and techniques hold the most promise for providing optimal educational and vocational guidance for students.

The results of evaluation of the vocational orientation sequence in the General College, University of Minnesota, reported in this paper, suggest certain possible answers to some of the questions posed above. It is the earnest hope of the author that they will serve to stimulate further research in this area to provide additional knowledge concerning the appropriate place and function of certain techniques in student personnel work.¹

¹ A report of the evaluation study was published by the Committee on Educational Research in 1941 (15). However, government restrictions on paper and other considerations limited the printing to a few hundred copies. The volume of requests for copies of the study has continually increased since exhaustion of the original supply several years ago. As a result, and at the urging of others, the author has submitted the report with minor revisions for publication in a professional journal of sufficient circulation to reach an appropriate audience.

Vocational Orientation in the General College

The courses in vocational orientation in the General College are the result of several years of development. In 1932, at the time the University Junior College² was established, a course in vocations was instituted as an elective. During the years 1936-38, *The Choice of an Occupation* was a required course for the *Psychology Comprehensive Examination* and a part of an Economic comprehensive which also included *Our Economic Life* and the *Mathematics of Business*. The present status of the vocational orientation area in the General College is an outcome of the 1938 proposal that four orientation areas, namely, Home Life, Individual, Social-Civic, and Vocational, be included in the curriculum.

As one of the orientation areas, the vocational orientation sequence was composed of three courses at the time of this study. The foundation course of the sequence, *The Choice of an Occupation*, was concerned with students' vocational choices. Because large numbers of students make obviously illogical and sometimes fantastic job selections, in terms of individual interests, aptitudes, abilities, and experiences, the course was aimed at developing an understanding of the problem of job selection which would create an awareness of the need for advice and help from competent sources. The materials presented to the students in this first quarter of work may be divided into four main sections: educational opportunities, human patterns, job patterns, and trends in employment. In the second quarter, *The World of Work*, occupational information was placed in a socio-economic setting and emphasis was given to relationships among workers and industries. To meet a need expressed by many students for a more individualized technique for exploring materials which fitted their particular vocational problems, the Vocational Orientation Laboratory was established. Whereas the first two quarters of work were taught primarily by classroom lectures, individual work and participation in small interest groups received first consideration in the laboratory. Students made a careful job analysis in the occupational field of their

² Later named the General College.

choice in addition to conducting specific investigations of job and training opportunities in the Twin City area (4, 5, 6).

Evaluation Objectives

Following the establishment of the four orientation areas, the General College asked the assistance of the University Committee on Educational Research in making an appraisal of the new program in vocational orientation. A subcommittee was appointed and work was begun under the chairmanship of Professor Donald G. Paterson of the Department of Psychology.³ A two-year project was planned, with the first year devoted to an exploration of the most suitable methods of investigation of the problem and the second year to the application of the methods developed in the first phase of the study.

The major objective of the study was to determine the extent to which the vocational orientation courses assisted students in making more appropriate vocational choices. In addition, it was proposed to evaluate the effectiveness with which certain other objectives of the course, as outlined by its sponsors, were being achieved, since findings from such a study would be of considerable significance to the problem of vocational counseling throughout the University. In particular it was hoped to determine the extent to which group counseling and instructional methods can (1) solve certain students' vocational problems, (2) sift out the students who need individual clinical guidance, and (3) reduce the time necessary for individual counseling after preliminary orientation.

Procedures Used in the Evaluation⁴

In conducting the investigation, tests and questionnaires were administered and ratings of vocational choices were made both at the beginning and at the end of the experimental period.

³ Members of the evaluation committee were: Malcolm S. McLean, T. Raymond McConnell, Edmund G. Williamson, John G. Darley, Ivor Spafford, Marcia Edwards, Ruth E. Eckert, Milton E. Hahn, C. Harold Stone, and Donald G. Paterson, chairman.

⁴ Since the work of the first year was primarily exploratory in nature, a detailing of research activities is not recorded here. The main organizational change resulting from the exploratory program was the establishment of the laboratory approach to the study of vocations at the beginning of the Winter Quarter, 1939.

The experiment was limited to members of the Freshman class, since the majority of students who elected to enroll in the vocational sequence did so in their first year. To allow adequate time for the testing, a day was set aside during the first week of the Fall Quarter for the pretest and another during the Spring Quarter for the retest. The entire Freshman class was given the battery of tests and questionnaires. This group was divided later into the students who actually took the vocational orientation sequence during the year and those who did not.

The following techniques and measures were used as the primary bases for evaluating the effectiveness of the vocational orientation core courses:

- A. Sources for obtaining general background information:
 1. General College Individual Record Form, from which was obtained such information as family background, educational history and planning, stated vocational choice and claimed interest, extracurricular activities, hobbies, etc
 2. The vocational autobiography: a 2,000-word paper written by the student covering present and past vocational choices and including a discussion of the factors that have influenced these choices, a work history, and plans for the future.
 3. A questionnaire designed to determine the amount of counseling a student had in high school and by whom he was counseled.
 4. Individual student case folders on file in the counseling office which contained complete records of all ability, aptitude, interest, and personality tests taken by the students, as well as information obtained by faculty advisers and counselors in personal interviews.
- B. Measures used in both the pretest and retest batteries:
 1. A comprehensive test of occupational information composed primarily of the most discriminating items of previous tests.
 2. An attitude scale adapted by Williamson and Darley from the *Occupational Attitude Scale* of Miller and Remmers covering seven general occupational fields.

- 3 Remmer's *Scale of Specific Attitudes* toward a Vocation to determine student attitudes toward their own vocational choice.
 4. A form developed by the Evaluation Committee to determine the training desires of students in various vocational and cultural course sequences.
 - 5 Preliminary form of the *Minnesota Personality Scale* to obtain measures of student adjustment
 - 6 An adjustment scale developed by the Individual Orientation Evaluation Committee to obtain additional evidence of adjustment or maladjustment among students.
- C Additional measures and techniques applied at time of the retest:
1. A special questionnaire to determine changes in attitude and thinking of students on vocational problems as expressed in the autobiography written at the beginning of the year. This form also sampled reactions to vocational counseling received during the year
 - 2 A ranking by students in order of preference, of all of the courses in which they were enrolled during the experimental period
 3. An evaluation of occupational choices in relation to claimed interest, high-school record, measured interests, and other test and clinical data An attempt was made to determine the consistency of the vocational choices a student listed at the beginning and the end of the year in relation to his measured abilities and capacities.
 4. A determination of student reactions to the vocational orientation courses. For this purpose, both objective and written responses were obtained from students at the end of each quarter.

Data obtained from measuring devices and evaluation techniques listed above, together with information drawn from individual case folders on file in the counseling offices, were analyzed in relation to counseling patterns, course sequences, sex, and reasonableness of vocational choice

Student Groups Included in the Study

In making comparisons between students in the control group and those taking vocational courses, experimental groups were established according to the course sequences chosen by various students. These are:

1. *Combined Experimental.* $N = 118$. This group included all students who were enrolled in vocational orientation courses during the year and took the complete pretest and retest batteries.
2. *Experimental I.* $N = 80$. Composed of students who took the *Fall and Winter Quarter lecture courses*, namely, *The Choice of an Occupation*, and *The World of Work*.
3. *Experimental II.* $N = 12$. Students enrolled only in the *first lecture course*, (*The Choice of an Occupation*) and in the *Vocational Laboratory*.
4. *Experimental III.* $N = 26$. Students who took the entire vocational orientation sequence, i.e., *both lecture courses and the Vocational Laboratory*.
5. *Control Group.* $N = 140$. All freshmen not enrolled in any of the vocational orientation courses and who took the complete pre- and post- test battery.

The control and experimental groups appear to be well matched on such factors as age, aptitude rating, high-school rank, measures of adjustment, and father's occupation. Of most immediate interest, perhaps, is the question of aptitude rating and high-school rank, since it is to be remembered that this group, although typical of the General College freshman population, is not typical when considered in relation to the total freshman population of the University of Minnesota. The mean aptitude test rating of the control group on the *American Council on Education Psychological Examination* is equivalent to a percentile score of 15 on all university of Minnesota freshman norms. For the combined experimental group the mean percentile is 19. The average high-school percentile rank of students in the control group is 34.5. The high-school rank for the combined experimental group is 33.8.

Results

Gains in Vocational Information

As one might expect, students who were enrolled in the vocational orientation sequence made greater gains in occupational information during the year than did students in the control group. At the beginning of the year, the mean scores for the combined experimental and for the control groups were almost identical. At the end of the year both groups showed gains. Residence in college for one year served to increase slightly the amount of information about jobs which students in the control group possessed. However, students in the experimental groups made significantly greater gains than those in the control. An analysis of covariance between the control and the combined experimental produced an F of 11.08 on the retest at the end of the year. The 1 per cent level of significance was 6.76. When final scores were adjusted for differences in initial scores, an F value of 17.07 was obtained, indicating that students in the experimental group gained significantly more than members of the control group. In the subgroups it was found that students who were enrolled in the laboratory section made somewhat larger gains than did those who were enrolled in the lecture sections only.

Levels of Aspiration

The levels of aspiration for students in various groups may be ascertained through consideration of a number of the measures used in this investigation. Expressions of attitudes toward occupations, salary expectations, and vocational choices classified on the Minnesota Scale of Occupations, contribute to an understanding of the types of vocational goals aspired to by these General College students.

Attitudes toward Occupations—In general, at the beginning of the year the students studied favored types of work involving executive responsibility, business contacts with people, artistic work, and technical or scientific work, in the order named. The type of work favored least was that which involved business detail. Differences on the general attitude scale from which

these results were obtained were statistically significant in but two instances and showed for the most part that the control and experimental groups were quite similar in their attitudes toward jobs.

By the end of the year certain changes in attitude had occurred. Only those changes which satisfied the 5 per cent level of significance are presented here ⁵ The combined experimental groups showed a much less favorable attitude toward occupations involving social service activities at the end of the year. Except for Experimental III (Fall and Winter lecture sequence plus laboratory) all groups were less favorable to occupations requiring special artistic abilities. In attitudes toward occupations involving executive responsibilities, only Experimental II showed a significant change, and this was in a direction less favorable to executive work. Technical or scientific work was considered much less desirable by students in both the control group and in the combined experimental group at the end of the year than they had rated it at the beginning of their college work. Students in the control group looked on occupations involving verbal or linguistic work with much less favor at the end of the year than at the beginning.

If the vocational orientation courses were assisting in the distribution of student vocational choices to levels more appropriate to their abilities one would rightfully expect to find students less favorable at the end of the year toward types of work demanding extended academic training. One demonstration of such an outcome is the change in attitude toward the social service occupations, most of which involve at least college graduation and often postgraduate work. Changes in attitude toward artistic occupations suggest that some factors in the General College were operating to cause all students to favor such work as a career less at the end of the year, and the vocational orientation courses operated to accentuate these changes. It is not clear what factors were operating to cause the change in attitude of the control group toward verbal and linguistic work without accompanying changes in the experimental group.

⁵ A critical ratio of 1.96 or more was considered as significant.

Salary Expectations—Changes in students' estimates of the monthly salaries they expected to receive their first work year, at the end of five years, and at the peak of their earning power, did not follow any consistent trends which might suggest the influence of the vocational orientation courses. The mean salary expectations for men students in all groups were approximately \$120 per month for the first year on the job, \$240 per month after five years of work, and \$450 per month for peak earnings. These estimates were made at the end of the experimental period and are not significantly different from estimates made at the beginning of the year.

Levels of Vocational Choice.—Occupational choices of male students and the occupations of their fathers were classified on the *Minnesota Occupational Scale* (3) in order to reveal whether the vocational orientation courses were performing the function of "downgrading" student vocational choices to more realistic levels. Student choices and fathers' occupations were compared also with occupations of employed males in the United States classified on the Minnesota scale by Smart (9) from the 1930 census. As might be expected in dealing with a college group, the occupations of students' parents showed a slightly larger percentage in the professions and in semiprofessional and managerial occupations than is true of Smart's general population group. In the *Minnesota Occupational Scale*, occupations are classified as follows.

Class	Occupation
I.	Professional
II.	Semiprofessional and managerial
III.	Clerical, skilled trades, and retail business
IV.	Farmers
V.	Semi-skilled, minor clerical, and minor business
VI.	Slightly skilled trades and other occupations requiring little training or ability
VII.	Day laborers of all classes

At the beginning of the experimental period students in the control group had choices which varied little from those of students in the experimental group except for Class II. A much larger percentage of control choices fell in this classification. As one might expect, both groups showed a rather large

percentage of choices in Class I. The relationships may be seen in Table 1

By the end of the year, students in the experimental groups appeared to have adjusted their vocational choices more readily to lower levels approximating the occupational levels of their parents than did students in the control group. Members of the control group showed an even less realistic pattern of choices at the end of the year than at the beginning and were much

TABLE 1

*Vocational Choices of Male Students at the Beginning and End of School Year Compared with Occupations of Their Fathers Classified on the Minnesota Occupational Scale**

	Control N = 13			Combined Experimental N = 74			Experimental I N = 46			Smart's Classification
	Student Choices		Father's Occupation	Student Choices		Father's Occupation	Student Choices		Father's Occupation	
	Be- ginning of Year	End of Year		Be- ginning of Year	End of Year		Be- ginning of Year	End of Year		
	%	%		%	%		%	%		
Class I. . . .	25.58	23.26	6.98	32.43	16.22	5.41	30.43	19.56	6.52	2.59
Class II	41.86	51.16	23.26	22.97	33.78	28.38	26.09	32.61	32.61	7.22
Class III. . .	23.26	11.63	32.56	31.08	20.27	28.38	32.61	23.92	28.26	13.81
Class IV.	—	2.32	4.65	1.35	1.35	4.05	2.17	2.17	2.17	15.42
Class V.	9.30	11.63	25.58	12.16	28.38	25.68	8.70	21.74	23.92	23.74
Class VI.	—	—	4.65	—	—	5.41	—	—	4.34	14.47
Class VII. . .	—	—	2.32	—	—	2.70	—	—	2.17	22.63

* Choices of female students are not included because of lack of comparability to father's occupation. Only those male students who had a definite choice at beginning and end of year are shown. Those who had no choice at either beginning or end are not shown.

farther away from the occupational classifications of their fathers.

Although Table 1 presents data for only the Combined Experimental Group and Experimental I in addition to the Control Group, the trends for the other experimental subgroups were all in the same direction, which would tend to lend further significance to the results shown for the combined group. *From the changes observed, it would appear that the vocational orientation courses were serving to downgrade the occupational aspirations of students to more realistic levels.*

Appropriateness of Vocational Choice

Although the downgrading of vocational choices of some students is to be regarded as a significant outcome, too much

credence should not be given, without further examination, to the appropriateness of these changes of choice. The change of a student's choice to a more appropriate level does not necessarily mean that he has selected a vocation that is in accord with his pattern of interests, abilities, and capacities.⁶ The suitability of the choice of each individual was therefore carefully rated by members of the General College counseling staff from data provided by individual case studies. Measures of appropriateness of choice were obtained by comparing counselors' ratings of students' vocational possibilities with stated vocational choices of the students at the beginning and end of the experimental period.

Accepting the premise that a particular pattern of abilities and interests allows an individual to cover a range of occupations within one or several broad fields of work, ratings were made in broad areas instead of indicating a specific occupation for each student. Going even further, counselors indicated not only one broad field of work in which each student appeared to have potentialities for successful adjustment, but in many cases listed as many as three alternatives. These alternatives were listed in the order of the counselor's judgment of the student's potentialities in each field. Within each broad area three levels were presented based on a hierarchy of the intelligence and training necessary for successful competition.

The broad areas on which judgments were made were ten in number and were based on T. L. Kelley's Classification of Human Abilities, including one general category listed as "practical" (10, 11). The ten fields in the ability classification of occupations used are listed below with a brief description of each:

1. Verbal and linguistic—fluency in use of one's own language and facility in learning other languages.
2. Scientific—facility in defining, classifying, grasping principles, inductive reasoning, perceiving relation of rule to example.
3. Mathematical—facility with abstract symbols (and relations of cause and effect). Perception of complex number relations.

⁶ However, even though students' individual choices may not be good, counseling is facilitated if alternatives fall at an acceptable level.

4. Clerical and commercial—accuracy and speed in handling numbers, names, systems, and details.

5. Constructive and mechanical—perception of spatial relations, facility in designing, calculating, working with machinery, etc.

6. Manual skills—dexterity in using tools, skills with hands and fingers, precision in coordinating movements.

7. Artistic—appreciation of form and color, facility in crafts and in imaginative interpretations.

8. Executive—initiative, self-reliance, ambition, leadership, etc.

9. Social—sociability, cooperativeness, tact, personal pleasingness, helpfulness.

10. Practical—efficiency in practical affairs, calmness under pressure, persistence, courage.

In making ratings of appropriateness of choice, four classifications were used: *Optimal*, where the student's choice coincided with both the field and level suggested by the counselor; *Fair*, where the student's choice was in the same field but at a different level from the counselor's rating; *Poor*, where neither field nor level coincided; and *No Choice*, where the student had made no vocational decision.⁷ In this report, only the optimal and poor choices have been graphed for descriptive purposes. In addition, no attention has been given to sex differences.

A comparison of changes in vocational choices of students in the control group with those of students in the experimental groups reveals relationships which have important implications for guidance programs of both the traditional group type and those utilizing the more comprehensive clinical method (12, 16).

When no consideration was given to the counseling received by students, the percentage of poor choices was not reduced by

⁷ To determine the consistency of counselors' ratings in this study three counselors made independent ratings of the same fifty cases. In comparing the occupational possibilities listed by the counselors, Counselors I and II had at least one agreement in each of the fifty cases (in many instances more than one vocational alternative was listed by a counselor). Counselors I and III agreed with the same high consistency—at least one agreement on occupational alternatives in every case. Counselors II and III had at least one agreement in forty-seven out of the fifty cases, or 94 per cent of the time.

the end of the year and the increase in proportion of optimal choices was insignificant. However, when the students who received vocational counseling during the year were separated from those who were not counseled, marked differences were

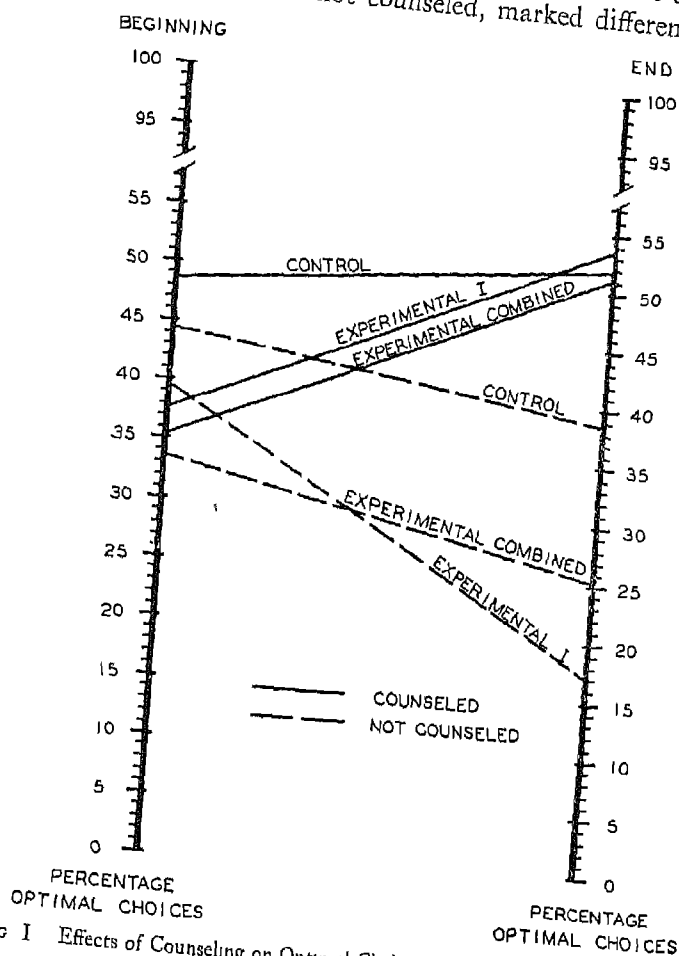


FIG I Effects of Counseling on Optimal Choices of Students in Different Groups

discovered. Figure I presents graphically the effects of counseling. The trend is the same for both control and experimental groups. The counseled students have a higher proportion of optimal choices at the end of the year and those not counseled have a smaller proportion of optimal choices. But, although the trends are the same, the effect in both directions

is far greater in the experimental groups. *Students who received instruction in vocational orientation in addition to individual counseling made marked gains in the proportion of optimal choices by the end of the year. Conversely, students who received instruction but were not counseled, showed much greater decrements in optimal choices than students in the control group who received neither counseling nor instruction.*

An analysis of the changes in percentage of poor choices during the year reveals additional trends which further challenge the use of courses in group guidance without providing adequate supplemental individual counseling. Students who received instruction in vocational orientation but were not counseled during the year were no better off than students in the control group who were not counseled. Figure II shows this relationship. It will be noted that the lines for non-counseled students in all three groups closely approximate one another. In the counseled group there is a reversal of trend. The combined experimental group and Experimental I both show a reduction of poor choices for the counseled students. Control students who received counseling show a slight increase in proportion of poor choices but the change from initial to final status is much less marked than is true for those who were not counseled.

Trends in the laboratory groups (Experimental II and III) were in the same direction as for the lecture group (Experimental I). However, fewer laboratory students who were not counseled had poor choices at the end of the year and there was a smaller decrement in optimal choices during the year. There was little difference in gains in optimal choice between the counseled laboratory and lecture groups. On the other hand, counseled laboratory students showed a marked reduction in proportion of poor choices by the end of the year.

From the general trend of the results, those most closely associated with the evaluation study judged that the second quarter lecture course dealing with information about jobs was not as effective in assisting students to better vocational choices as the first quarter of the sequence which dealt with individual differences and specific steps to follow in choosing an occupation. Where occupational information was directed

to the individual as in the laboratory approach, it appeared to be more effective.

Other analyses were made to discover differential factors in student vocational choices. Students who changed their

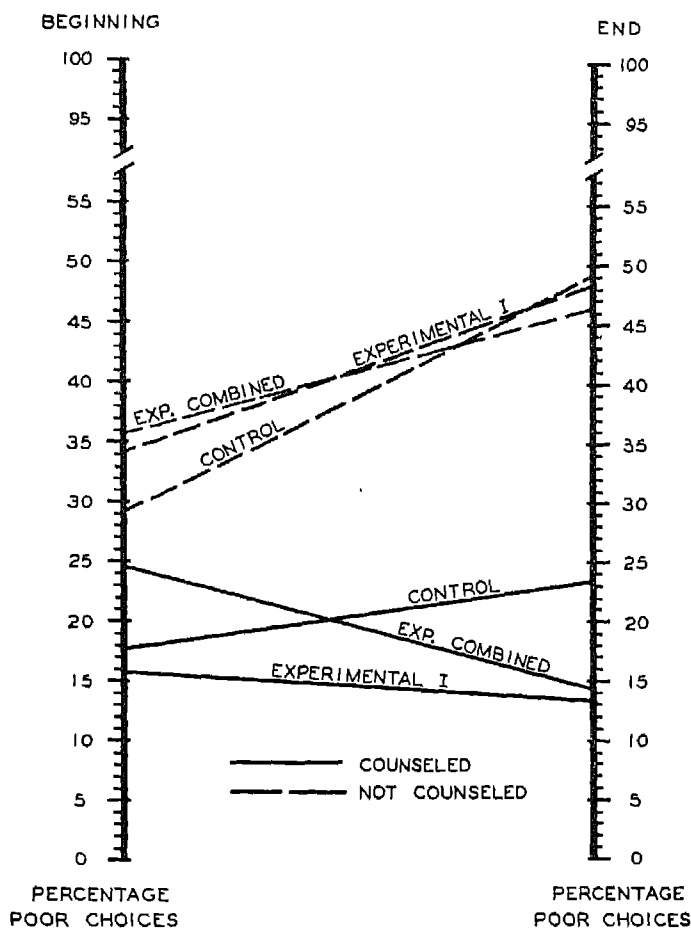


FIG. II. Effects of Counseling on Poor Choices of Students in Different Groups

choice during the course of the year were separated from those who had maintained the same choice. When these groups were considered in relation to counseling, it was found (as would be expected) that students who *had made changes in choice* made them in the same directions as shown in Figures I and II, and the change from initial to final status in each group was much

greater. For students who had made *no changes* in choice, counseling again appeared as an important factor. In all groups, a much larger percentage of students who were counseled maintained good choices than was true of the noncounseled groups. Conversely, a very small percentage of students in the counseled groups persisted in poor choices during the year. In the noncounseled groups a much larger proportion of students expressed the same inappropriate choices at the end of the year that they had at the beginning. *The vocational orientation course supplemented by counseling was much more effective in causing students with poor choices to change them than was counseling alone.*

Reactions of Staff Counselors

Since these results reveal definite values from the combination of vocational courses and counseling, it appeared likely that statements from the counselors who dealt with the students might provide additional information concerning the way in which such a combination actually functioned in individual cases. Both of the full-time staff counselors therefore were asked to indicate whether or not differences had been observed in the counseling situation between students enrolled in the course and those who were not. The counselors' statements (prepared independently) are given below:

Royal B. Embree, Jr., instructor and research counselor:

It is impossible to provide an objective or quantitative estimate of the value of vocational orientation courses to the individual counseling activity carried on by General College counselors. A subjective review of experience with many interviews would lead me to believe that these courses are of great value to the counselor. Classes in vocational orientation have furnished good initial orientation to the various measuring devices used in General College, as well as a sound foundation upon which vocational counseling interviews can be organized. Vocational autobiographies, supplied by these courses, are valuable additions to the personnel records. Finally, vocational orientation courses have been a definite force in motivating students to come for counseling interviews.

Cornelia Taylor Williams, instructor and research counselor:

It is my distinct impression that students who had had, or were taking, the vocational orientation course were easier to

advise on vocational and educational problems than many of those who had not had the course. Students who came in for vocational advice often told me that they needed to know something about their own abilities and limitations. They seemed to have enough preliminary and basic understanding of "aptitude testing" so that explanation and interpretation to them of their own test scores was easier than for students who had not taken vocational orientation. They were, in other words, "softened up" to the point where they were ready for counseling, and could see some value in it, and they were well enough prepared for the interview so that the test interpretations proceeded more smoothly. They asked more intelligent and penetrating questions, and they understood better what I was trying to tell them.

It would appear then that the vocational orientation courses served as a preparation for counseling and tended to reduce the amount of time necessary for the resolution of vocational and educational problems.

Measures of Adjustment

The achievement of more adequate social and emotional adjustment has been suggested as a significant outcome of vocational orientation courses. For this reason, two rather comprehensive measures of adjustment were employed. The first, a preliminary form of the *Minnesota Personality Scale* (1), provided measures of adjustment in five areas, namely:

- I. Morale
- II. Social adjustment
- III. Family relations
- IV. Emotionality
- V. Economic conservatism

The second, the *Individual Orientation Adjectives Test* (2) which was developed in connection with the evaluation of the individual orientation area, provided four types of scores which seemed useful for this purpose. These four categories are:

- I. General characterization or pattern of individual's total behavior
- II. Typical behavior as a result of motivation
- III. Characteristic feelings of individual
- IV. Social characteristics of individual

On the *Minnesota Personality Scale* the control group made no significant changes on any of the five factors covered. The experimental groups registered significant changes only in the area of social adjustment. An investigation of the effect of counseling on adjustment revealed an interesting outcome. Students in the combined experimental group who were not counseled showed no significant changes on any of the measures in this test. However, students who were counseled showed significantly better social adjustment at the end of the year, as measured by this test. Similar breakdowns for the control group revealed no significant changes.

On all four divisions of the *Individual Orientation Adjectives Test* only very slight changes were recorded in all groups. These same results held true when groups were divided according to counseling received and according to course sequences in which students were enrolled.

These results suggest that student adjustment does not show very marked changes as a result of one year's residence in the General College. However, *a combination of counseling and course instruction in vocational orientation does produce significant changes in the social adjustment of individuals as measured by the Minnesota Personality Scale.*

Summary and Conclusions

During the school years of 1938-39 and 1939-40 the Committee on Educational Research of the University of Minnesota conducted an evaluation study of the core courses in the Vocational Orientation Area in the curriculum of the General College. The first year of the investigation was exploratory in nature. The results described in this report were obtained during the second year. The vocational orientation sequence included two lecture courses: Choice of an Occupation, and The World of Work, and Laboratory sections in which individual work was emphasized. Freshmen who were enrolled in the vocational orientation courses formed the experimental groups and the balance of the Freshman class served as the control. A total of 258 students had records which were complete enough to warrant their inclusion in the study. The combined experimental group numbered 118 students and there

were 140 students in the control. In order to determine changes which might be attributed to the teaching of vocational orientation, a variety of measures was administered at the beginning and end of the year to all members of the Freshman class. Studies were made of changes in occupational information, attitudes toward occupations, salary expectations, levels of vocational choice, suitability of vocational choice, direction of changes of vocational choice, and social and emotional adjustment.

An analysis of results suggest the following conclusions:

1. The vocational orientation courses in the General College provided students with more information about jobs than students gained outside of the course through undirected and incidental reading.

2. Changes in attitudes towards occupations in both the control and combined experimental groups reflected less favorable attitudes at the end of the year toward technical and scientific work and occupations involving artistic abilities. The combined experimental group showed significantly less favorable attitudes toward social service occupations which normally require extended academic training. Students in Experimental II (course sequence: Choice of an Occupation and Vocational Laboratory) were less favorable toward work involving executive responsibilities, and control students were less favorable toward occupations involving verbal or linguistic work.

3. The vocational orientation courses did not bring about any changes in the average salary expectations of students which would not normally arise from a year's residence in the General College.

4. Students in the experimental groups tended to adjust their vocational choices to more appropriate levels than students in the control group.

5. Vocational orientation courses alone did not cause students to make more appropriate vocational choices in terms of their abilities, aptitudes, and interests than did mere residence in college.

6. Students enrolled in vocational orientation courses who were counseled made the largest gain in proportion of optimal

choices from the beginning to the end of the year, whereas students enrolled in vocational orientation courses who did not receive counseling showed the greatest decrement in proportion of optimal choices from the beginning to the end of the year.

7. The vocational orientation courses alone did not serve to offset the tendency of uncounseled students to make a greater proportion of poor choices at the end of one year's residence in college. However, when students in these courses were counseled they showed a reduction in the proportion of poor choices by the end of the year.

8. Vocational orientation served as a preparation for counseling and tended to reduce the amount of time necessary for the resolution of vocational and educational problems.

9. Although no other changes in measures of adjustment were observed, significant changes in social adjustment in a favorable direction resulted from a combination of counseling and course instruction in vocational orientation

The general consistency of the trends of this investigation suggests the urgent need for a careful re-examination of programs of vocational and educational guidance as they exist in both secondary schools and higher institutions. In many instances educators have felt that the mere process of teaching about occupations and allowing students to make judgments of their abilities through self-analysis serves as ample means to aid students in making appropriate vocational choices. On the other hand, some educators have given all of their attention to a careful and objective analysis of the individual with no time left for thoughtful consideration of occupational opportunities in fields of work which are in accord with the abilities of the student. The trends of this study indicate that the adequacy of either method of guidance when used alone is open to serious question. Thoughtful attention to the objective analysis of individual aptitudes, abilities, interests, personality characteristics, and background, and the presentation of techniques for making careful choices, together with the presentation of vocational information and discussions of current occupational opportunities, promise the most successful assistance to students in preparing for their work and entering upon it.

REFERENCES

1. Darley, J. G. and McNamara, W. J. *Manual of Directions, Minnesota Personality Scale*. New York: Psychological Corporation, 1941.
2. Fattu, N. "Evaluation Program in Individual Orientation," *Studies in Higher Education*. Minneapolis: University of Minnesota, 1941. pp. 108-130.
3. Goodenough, Florence and Anderson, John E. *Experimental Child Study*. New York: Century Company, 1931.
4. Hahn, M. E. "Vocational Orientation," *Curriculum Making in the General College*. Minneapolis: University of Minnesota Mimeograph Department, June, 1940. pp. 138-56.
5. ———, "Vocational Orientation." *Journal of Higher Education*, II (1940), 233-237.
6. ———, "Occupational Orientation for College Students." *Occupations*, XVIII (1940), 432-435.
7. Hahn, Milton E. and Brayfield, Arthur H. *Job Exploration Workbook*. Chicago, Illinois: Science Research Associates, 1945.
8. ——— — ———, *Occupational Laboratory Manual*. Chicago, Illinois: Science Research Associates, 1945.
9. Institute of Child Welfare, University of Minnesota. "Classification of Occupations of Employed Males in the United States." *Alphabetical Occupational List for Classification of Occupations by Minnesota Occupational Scale*, (Mimeographed). p. 8.
10. Kelly, T. L. *Cross-Roads in the Mind of Man*. California: Stanford University Press, 1928.
11. ———, *Essential Mental Traits*. Cambridge, Massachusetts: Harvard University Press, 1935.
12. Paterson, D. G., Schneidler, Gwendolyn G., and Williamson, E. G. *Student Guidance Techniques*. New York: McGraw-Hill Book Company, 1938.
13. Remmers, H. H. et al. "Studies in Attitudes." *Bulletin of Purdue University*, XXXV (1934), 1-112.
14. ———, "Further Studies in Attitudes, Series II." *Bulletin of Purdue University*, XXXVII (1936), 1-298.
15. Stone, C. H. "Evaluation Program in Vocational Orientation." *Studies in Higher Education*. Minneapolis: University of Minnesota, 1941.
16. Williamson, E. G. and Darley, J. G. *Student Personnel Work*. New York: McGraw-Hill Book Company, 1937.

AN OPTOMETRIC APTITUDE TEST

NEIL D. WARREN AND A. A. CANFIELD

University of Southern California

As a basis for the construction of an aptitude test for training in optometry, certain preliminary steps were necessary. A review of the literature in the fields of medicine and engineering revealed that the tests for training in medicine stressed the verbal achievement of the applicant, and that those in engineering stressed the mathematical achievement. Inasmuch as optometry seems to combine studies related to both of these fields it was thought logical to make some measurement of these two factors in the aptitude test for optometry. In addition to the information gained from a review of the literature in related fields, a job analysis was made on practicing optometrists in the city of Los Angeles. The ability to read dials and tables accurately and quickly and the ability to detect diseased conditions of the human eye seemed to be two important components of their duties. A section of the test was devoted to an attempt to measure the applicant's ability to read a graph quickly and accurately and to note minor variations in detail. In addition to the review of the literature and the job analysis, a careful examination of the curriculum of study at an optometry school was made. The school administration pointed out those subjects that were of most importance to the student and special emphasis was given to them. The resulting test consisted of five parts and was to determine the following characteristics of the applicants:

- A. General level of scientific information.
- B. Scientific interest and attitude.
- C. Certain personality factors, such as attention to detail, accuracy of work, speed of work, etc.
- D. Amount of general information in fields other than science.
- E. Mathematical ability

F. Ability to do fine work which demands close attention and accurate perception.

The primary purpose of the test was to determine whether or not prospective students would be able to assimilate the material offered in an optometry school. It was not the purpose of the test to select those applicants who would do well in the practice of optometry or who would demonstrate rigid professional ethics in the future, "except insofar as accomplishment in school is related to professional success.

Description of the Test

Part I. Figure Checking.—The first part of the test was composed of items of the familiar number checking type, designed to measure visual acuity, speed of perception, and persistent attention to detail. Four and one-half minutes were allowed to complete 40 items similar to the sample below:

A. 147316598.

B. \$%o(#&)(?D.

147315689

\$%o(#&())?D#

The student was to mark A on his answer sheet if both sets of figures were the same, and B if they were different in any way.

Part II. Personal History—The second part was composed of 50 items, dealing with the interest, attitudes, and personal history of the student. Eighteen minutes were allowed for this part of the test. The interest items were designed to measure the relative amounts of scientific and persuasive interest. Sample items appear below:

C. In general glasses are:

- a. attractive.
- b. barely noticeable.
- c. unsightly.

D. Of the following things, I would prefer to

- a. clean laboratory apparatus.
- b. clerk in a store.
- c. do rough carpentry.
- d. tune pianos.

Part III. Mathematics.—Forty mathematical problems were selected to cover algebra, trigonometry, and mathematical

reasoning. The students were allowed 45 minutes to solve them. A sample item appears below:

- E. A is twice as old as B. C is $\frac{1}{2}$ as old as B.
How many times older than C is A?
- a. 2
 - b. 3
 - c. 4
 - d. 6
 - e. 8

Part IV. Verbal Reasoning and Vocabulary.—The fourth part of the test consisted of 80 items designed to measure general mental ability and scientific knowledge and 27 minutes were allowed for its completion. General information items covering a wide range of topics, primarily from the biological and physiological sciences, were used along with science vocabulary items and reasoning items. A sample item follows:

- F. Of the following, which has the most highly developed nervous system?
- a. Hydra
 - b. Spiders
 - c. Worms
 - d. Beetles
 - e. Amoeba

Part V. Graph Reading.—The fifth part of the test was a complicated graph reading task consisting of 25 items. The student was required to find a point, interpolate its position and move it across a graph according to a set of instructions. This test was designed to measure attention to detail, thoroughness of work, and persistence of attention. The students were given 20 minutes to complete all of the items.

Procedure

The test was given to two classes for validation purposes. The first study was made on the class of 1947, during their senior year in the school. Sixty-five students participated in this study, and constitute what will be referred to hereafter as Sample I. The second study was conducted in the early months of 1947 on the class of 1948, during their junior year in the school. Eighty-two students took the test in this second

study. The 82 students who constituted the second validation sample will be referred to hereafter as Sample II. All of these students, the 65 in Sample I and the 82 in Sample II, had completed at least two years of college training elsewhere prior to their entrance into the Los Angeles School of Optometry.

Each section of the test had a time limit, the time allowed for Parts II, III, and IV being long enough to permit most of the students to finish those sections of the test. Parts I and V were speed tests, and hence the time-limit was set so short that only a very few would finish in the time allowed. Table 1 shows the mean score and the standard deviation of scores for both samples for each part of the test and for a total score,

TABLE 1
The Mean and Standard Deviation of Total and Part Test Scores

	Sample I N 65		Sample II N=82	
	Mean	SD	Mean	SD
Part I (Number Checking)	26.14	7.04	28.50	6.06
Part II (Personal History)	1.28	3.96	.61	2.72
Part III (Mathematics)	10.48	5.22	11.35	4.64
Part IV (Science Information)	38.09	11.00	40.18	10.26
Part V (Graph Reading)	12.56	6.54	10.92	7.62
Total (I, III, IV and V)	87.20	24.28	90.50	21.55

based on a summation of scores obtained on Parts I, III, IV, and V.

The means and standard deviations reported for Part II of the test (Personal History, Interest, and Attitude items) were obtained in the following manner. No attempt was made to formulate a key on these items on a priori grounds. A key was developed by means of an item analysis made on the papers of the upper and lower 27 per cent of the students in the first sample. This key was then applied to the total groups in both Samples I and II. Since some of the items had responses that were both favorable and unfavorable to success in school, some items were negatively weighted and hence it was possible for an examinee to obtain negative scores on this section of the test.

The scores made on the various parts of the test, and a total

score based on a simple summation of scores on the sections of the test, excluding Part II, were then correlated with the students' scholastic averages in school. Part II scores were not considered in deriving a total score because it was not believed statistically sound to use them. The fact that the key was prepared on some of the group and then applied to the total group would in itself insure spurious validity. Also, the number of items in Part II that had Phi's of significance was very small and the part was scored more as a preliminary investigation than as a predictive index.

The school used the letter system for grading and graded from A to F. Since such grades as A-, C-, etc. were reported, they were converted into numbers for ease of manipulation. The average grade made during the first year of study at the optometry school, (the junior year in college), was used as the criterion of success in training. The study was extended to include prediction of senior grades and a two-year average grade, but since the elimination or termination rate is so high the first year as contrasted with the second and last, the junior grades seemed to be an adequate and logical criterion to use in evaluating the effectiveness of the test.

Results

All correlations were computed by the Pearson product-moment method and the usual assumptions of the method seemed to be met. All results on all parts of the test seemed to indicate a normal distribution of scores, and linearity of regression was observed. The small samples used in the study introduced fluctuations in the normality of distributions that could hardly be classed as error variance. In order to combine the two samples into one, and to make a combined study, their raw scores on the test and on the criterion were converted to C-scores. The reduction of the number of intervals allowed in grouping the data may have reduced the correlations slightly, but no correction is normally applied if the number of intervals is greater than ten. The correlations obtained between the students' average junior grades and their scores on the various parts of the aptitude test are presented in Table 2

Parts III, IV, and V of the test maintained their validities quite well from the first sample to the second; however, both Parts I and II had noticeable drops. The sudden drop in the validity of Part II is understandable in terms of the method by which the first correlation was obtained, and indicates that the items selected for the first study had no validity when applied to another. Part I, the Number Checking section, dropped noticeably also. No change in the curriculum occurred between these two validations, and no explanation of this change of correlation can be offered other than a chance fluctuation produced by small sample statistics.

TABLE 2
Correlations between Average Junior Grades and Scores on the Optometric Aptitude Test

	Sample I N 65		Sample II N 82		Samples I & II N=147	
	r	SE	r	SE	r	SE
I (Number Checking)	.42	.10	.24	.10	.28	.08
II (Personal History)59	.08	.10	.11	—	—
III (Mathematics)48	.09	.55	.08	.48	.06
IV (Science Information)58	.08	.55	.08	.53	.06
V (Graph Reading)42	.10	.48	.09	.48	.06
Total (I, III, IV, V)62	.08	.59	.07	.57	.06

The solution of a multiple correlation equation on the two sets of data reveals more markedly the drop in effectiveness of Part I of the test. The "b" weights, derived by the Doolittle method, for the various Parts for the two administrations were as follows: Part I had a "b" weight of .1734 for the first validation data and a $-.2132$ for the second, Part III had "b" weights for the same two samples of .1878 and .7478, Part IV was .3967 and .3086, and Part V had "b" weights for the two samples of .2649 and .4048. The nature of the derivation of Beta weights, upon which the "b" weights are based, takes advantage of every minor variation in the regressions of the independent variables, and hence fluctuation would be expected to occur when computing the regression equation on groups of this size. No effort was made to compute "b" weights for Part II inasmuch as the validity coefficients themselves for that section of the test had no reliably significant meaning.

The reliability of the various parts of the test was computed on the basis of odd-even item scores and then was corrected by the Spearman-Brown prophecy formula. The reliabilities of the parts were in close agreement for the two administrations, being .94 and .94 for Part I, .82 and .84 for Part III, .83 and .84 for Part IV, and .95 and .99 for Part V. Since all parts of the test were timed, and Parts I and V were heavily weighted with the "speed" component, the reliabilities obtained by this method of computation are somewhat spurious. However, it is believed that sufficient reliability has been demonstrated. The inter-correlation of the various parts was computed by the product-moment method and yielded results, based upon the combined samples, as presented in Table 3

TABLE 3
Inter-correlations of Various Parts of the Optometric Aptitude Test

Test Part*	N=147					
	III		IV		V	
	<i>r</i>	<i>SE</i>	<i>r</i>	<i>SE</i>	<i>r</i>	<i>SE</i>
I						
III	.37	.07	.42	.07	.38	.07
IV			.55	.06	.42	.07
					.41	.07

* Part I—Number Checking, Part III—Mathematics, Part IV—Science Information, Part V—Graph Reading.

The relatively high inter-correlation of the scores on Parts III (Mathematical Ability) and Part IV (General Science Information) is probably due to the inclusion in both sections of items that involve reasoning and a certain amount of verbal comprehension. The relatively high inter-correlation between Part IV and all the other parts of the test indicates that verbal ability or some reasoning factor must be present in all parts of the test. The generally higher correlations obtained with Part IV of the test and school grades indicates that this section deals with those factors most closely related to the attainment of satisfactory grades in optometric training.

As a further check on the validity of the test, senior grades and an average two-year grade were obtained for all 65 of the students in Sample I. The predictive efficiency of the total

score, based on a simple summation of scores made on Parts I, III, IV, and V of the test, maintains itself well, being .54 for senior grades and .60 for the two-year average grade, which of course includes the junior average. The validity for Junior grades for this sample was .62. Part II (Biographical Information, Interest and Attitude) of the test maintained its validity at about the same level, giving correlations of .56 for senior grade average and .61 for the two-year scholastic average. Part III (Mathematics) correlated much lower with senior grades than junior, being only .29 for senior grades and .53 for the two year average grade. The other sections of the test did not vary in their validities more than .06. This is well within the margin of error.

It would have been advisable to administer a standard intelligence examination to both groups of students used in this study, and to correlate their scores on those standardized examinations with their grades in school. This would have been a good check on the usefulness of the *Optometric Aptitude Test* contrasted with that of a paper and pencil intelligence examination. Because of the limited time available for the study at the school, this procedure was not possible. However, the *Optometric Aptitude Test* and the Henmon-Nelson were given to 45 students in Elementary Experimental Psychology at the University of Southern California. The total score on the *Optometric Aptitude Test* and total score on the Henmon-Nelson correlated .43. Since the Henmon-Nelson is composed of three relatively unique types of items, Vocabulary, Arithmetic, and Reasoning, three scores were isolated on this test for each student. Correlation between various parts of the *Optometric Aptitude Test* and these rather artificial part scores on the Henmon-Nelson were computed. The verbal items from the Henmon-Nelson and Part IV of the *Optometric Aptitude Test* correlated .65, the reasoning items and Part IV correlated .70. The number items and Part III of the *Optometric Aptitude Test* correlated .59, and Part IV and the total score on the Henmon-Nelson correlated .61. It appears that the *Optometric Aptitude Test* is more than an intelligence test, combining some of those elements with achievement measures and clerical ability measures not usually measured as such in in-

telligence examinations. It is predicted by the authors that the *Optometric Aptitude Test* would have more validity for the selection of students in this field than standard paper and pencil intelligence examinations because of this wider coverage of ability and achievement

Conclusions

1. The *Optometric Aptitude Test* has demonstrated a validity on two samples of optometry school students that is very near .60. Under the circumstances, involving the many sources of variance that influence scholastic achievement, this is thought to represent a fairly strong correlation

2. A general science knowledge seems to be most closely related to success in this type of training compared to the other types of abilities and skills sampled by this test.

3. Biographical Information, Interest, and Attitude items, as constructed and applied in this test do not have consistent validity in predicting school grades from sample to sample, but do maintain their predictive efficiency in predicting senior grades from junior grades. The high correlation between first- and second-year grades (.77) no doubt largely accounts for this

4. Number Checking tests, such as the one included in the *Optometric Aptitude Test*, have validity for selecting optometry students, as do tests that involve an ability to read graphs quickly and accurately.

5. The ability to do mathematical problems involving reasoning are indicative of future success in the study of optometry.

The test seems to have sufficient validity to warrant its use by Optometry schools in student selection programs. The authors suggest that it be used with other indices of student promise such as past school and social records, recommendations by competent and ethical members of the profession, and personal interviews aimed at discovering the students' attitude toward the profession

SELECTION AND EVALUATION OF WEST POINT CADETS¹

DONALD E. BAIER
Personnel Research Section, A. G. O.

Introduction

THIS paper will sketch the major work by the Personnel Research Section of The Adjutant General's Office on the problems of selecting and evaluating West Point cadets. It is desired to make clear at the outset that the studies to be discussed have required participation by various staff members over a number of years. While the author shares general responsibility, it is primarily as a spokesman for some of the accomplishments of long-range institutionalized research exemplified by the West Point studies that he presents what follows.

West Point Qualifying Examinations

The first work done by the Personnel Research Section began in the spring of 1942 with the problem of selecting, from among West Point applicants with a record of successful college attendance, those who could be expected to fail scholastically at the Military Academy. In 1942 more than half of all of the applicants admitted to West Point qualified without examination by presenting a certificate of satisfactory educational standing from an accredited university, college, or technical school. A disproportionately large percentage of academic failures at the Academy came from this group, apparently because of the difficulties inherent in the evaluation of college records. To meet the need for identification of potential failures from this group and to provide a means of simplifying examining procedures for other modes of gaining admission, the construction of a scholastic aptitude examination was undertaken.

¹ This paper presents the personal views of the author and does not necessarily represent the official views of the Department of the Army

An analysis of the causes of academic failure indicated that Mathematics, English and Foreign Language, in that order, were the principal contributors to dismissal for scholastic failure, most of the attrition occurring during the first year. Accordingly, three tests covering basic mathematical concepts, arithmetic reasoning, and language aptitude were constructed to form, with the *Army General Classification Test*, a trial battery of machine scorable tests.

The *Elementary Mathematics Test*, a 30-minute test of 75 four-choice items, was constructed to get at mathematical ingenuity. The problems can be solved quickly if a simple algebraic or computational equivalence is recognized, but require tedious computation if solved step by step.

TABLE 1
*Correlations of West Point Qualifying Examinations with First Term Grade
for West Point Class of 1945*

Course	GC I	Elem Math	Lang. Apt	Ar Reas
Mathematics...	.43	.65	.50	.49
Spanish	.19	.30	.60	.40
English	.40	.44	.55	.19
Military Top	.40	.47	.30	.38
Tactics	.29	.30	.33	.32

The *Arithmetic Reasoning Test*, consisting of 54 four-choice items with a 30-minute time limit, involved no unusual features.

The *Language Aptitude Test* requires the translation of English into an artificial language and vice versa. The artificial language is described in a basic vocabulary and with rules for interchanging word forms. It assumes the grammatical rules of English. Given with a 30-minute time limit, the test consists of 59 four-choice items.

Scores resulting from the application of this battery to 932 cadets accepted in 1942 into the class of 1945, were correlated with first-term grades in Mathematics, Spanish, English, Military Topography, and Tactics. These are shown in Table 1.

After a consideration of the test inter-correlations and multiples it was decided that the most satisfactory battery would consist of the *Elementary Mathematics Test* and *The Language Aptitude Test*. This combination was titled the *West Point Qualifying Examination I*, hereafter referred to as WPQ-I. The

correlation between test scores and grades in succeeding years for the class of 1945 is shown in Table 2.

The decrease in the r 's in succeeding years is attributed to the progressive selection of the population with a consequent elimination in the lower range of scores due to academic failure and separation from the academy.

For WPQ-2 applied to the class of 1946 ($N = 826$), the sub-test vs. course correlations are comparable to those shown in Table 2. The relation between *total academic grade* for the

TABLE 2
Correlations of WPQ-1 Sub-Test Scores with Successive Course Grades at West Point for Class of 1945

WPQ-1 Sub-Test	Course	r for test vs. grades for			
		1st Semester	1st Year	2d Year	3d Year
Mathematics	Mathematics	.65	.60	.45	44
Language	English	.55	.54		41
Language	French		.46	.36	35
Language	German		.63	.52	48
Language	Spanish	.60	.44	.35	35
Language	Portuguese		.53	.27	27

entering year and *total WPQ score* is represented by an r of .54. For the graduating year it is .49.

That the most recent form of the *West Point Qualifying Examination* WPQ-3 is even more predictive than earlier forms is indicated in Table 3 which presents a comparison of WPQ-1 and 3 based on the class of 1947 which took both tests. The attenuation of the correlations in successive terms is again apparent.

Data derived from previous studies, not reported herein, supported the conclusion that the total score on the *West Point Qualifying Examination* predicts individual course grades as well as the appropriate sub-test score. Data from the class of 1947 on both WPQ-1 and WPQ-3 lend further support to this conclusion. They are presented in Table 4. Because of the administrative convenience of using a single score, this characteristic of the *West Point Qualifying Examination* added to its substantial validity, shortness, objectivity, and ease of administration and scoring has made the test an efficient element in the selection program for West Point cadets.

Obviously, there are other characteristics of the applicants for admission to West Point besides scholastic aptitude which deserve careful consideration in making a decision as to their suitability. One of these is aptitude or ability in the performance of physical activities as distinct from the relatively static basis for evaluating physical capacity derived from the medical examination. The Personnel Research Section has been ap-

TABLE 3
Correlation of WPQ-1 and WPQ-3 Subtest Scores with Successive Course Grades at West Point for Class of 1947

WPQ Subtest	Course	<i>r</i> for WPQ-1 Test Scores vs. Course Grades for		<i>r</i> for WPQ-3 Test Scores vs. Course Grades for	
		1st Term	1st Year	1st Term	1st Year
Mathematics	Mathematics	.74	.63	.79	.71
Language	English	.63	.56	.64	.61
Language	French	.38	.42	.44	.48
Language	German	.44	.47	.44	.45
Language	Spanish	.68	.67	.72	.73
Language	Portuguese	.51	.44	.56	.55

TABLE 4
Correlations of WPQ-1 and WPQ-3 Test Scores with First Semester Course Grades for Class of 1947

Course	WPQ-1 Test Scores			WPQ-3 Test Scores		
	Math.	Lang.	Total	Math.	Lang.	Total
Mathematics . .	.74	.54	.74	.79	.57	.79
English50	.63	.65	.52	.64	.65
French25	.38	.39	.22	.44	.37
German32	.44	.44	.37	.44	.46
Spanish49	.68	.65	.46	.72	.66
Portuguese25	.51	.44	.28	.56	.47

plying measurement techniques to this problem for more than a year.

(Parenthetically, it may be of interest to know that while the prediction of scholastic success has been under study since 1942, other aspects of the problem of selecting and evaluating the cadets of the Military Academy have been the subject of a long-range research program only since the Spring of 1946. The research to be outlined hereafter was requested at that time by the Superintendent of the Academy, Major General Maxwell D. Taylor, and initiated by the Adjutant General's Office immediately thereafter.) Up to the present time the following

has been accomplished. The Wherry-Doolittle test-selection method has been applied to some 19 physical proficiency tests including the Vertical Jump, Chins, Dodge Run, Sit-up speed, Softball throw, 300 yard Run, Dips, Standing Broad Jump, Rope climb, Vault, Burpee, Total Sit-Ups, Medicine Ball Put, Push-ups, 3 Broad Jumps, Flexion, Extensions, 50 yard Dash, and Endurance Index. Two criteria were employed for this purpose; one was grades in Physical Training, the other, an index of physical proficiency based on inter-cadet ratings. On the basis of these data, which will not be presented in detail because of the limited time available, a shortened, but efficient predictor battery consisting of 4 tests (Chins, Softball Throw, Standing Broad Jump, and Vault) yielding a shrunken multiple of .80 with course grades was employed for examining applicants in the spring of 1947.

The biserial correlation between a ten-test predictor battery and overall success or failure at West Point proved to be .13 although tetrachoric r 's at various points of cut indicated a higher value (.23) for the lower end of the distribution of scores.

Factor analysis studies are in process aimed at identifying the principal factors involved in physical proficiency measurement in order that further improvement in test batteries may be effected. Exploratory studies are also being planned to determine if possible the relationship of physical performance as a cadet to later officer performance.

Other traits of great importance to cadet selection, if they can be successfully measured, are those related to adjustment at the academy and later success as a regular army officer. For this purpose an experimental, multiple response, biographical and personality inventory was administered to the applicants for admission to the class of 1951. A long-term follow-up is contemplated to collect criterion data indicative of adjustment.

The construction of criteria for research purposes and the evaluation of cadet performance pose closely related problems. The principal research efforts in this area thus far have aimed at:

1. Studying the inter-relations among three of the major elements which determine final cadet standing. These are academic grade, physical education grade, and aptitude for the service rating. The last named is a composite score derived

from inter-cadet ratings and officer ratings by academic, tactical and physical training officers.

2. Determining the validity of the aptitude for service ratings, as well as other predictors, against success as an army officer as indicated by the Officers' Efficiency Rating.

Typical of inter-correlations among the major elements in final cadet standing are those for the graduating year, Class of 1946, shown in Table 5. Particular attention is directed to the high correlation between final year standing and final year academic grade.

TABLE 5
Intercorrelations of Major Elements and Final Standing for the Graduating Year, West Point Class of 1946 (N=770)

	1	2	3	4
1. Physical Ed. Grade		.28	.02	.06
2. Apt. for Service	.28		.29	.41
3. Acad. Grade	.02	.29		.98
4. Final Year's Stand.	.06	.41	.98	

TABLE 6
Intercorrelations of Semi-annual Officer Efficiency Ratings 18 Months after Graduation for West Point Class of 1944 and Variables Indicated. (N=403)

	1	2	3	4
1. Physical Ed. Grade (Grad. Year)		.39	.13	.17
2. Apt. for Service (Grad. Year)	.39		.46	.34
3. Final Composite Standing	.13	.46		.06
4. Officer Efficiency Rating	.17	.34	.06	

The relationship of the variables included in Table 5 to success as an army officer 18 months after graduating is shown in Table 6 for the Class of 1944. Officers' success is measured by score on the semi-annual efficiency report. The number of cases (403) is less than the total commissioned at the time of graduation, by the number of officers for whom no semi-annual efficiency rating could be obtained. In evaluating the data presented in Tables 5 and 6 the official weights assigned to the elements should be borne in mind. From a total of 1220 points

Academic receives 890

Aptitude for Ser. 90

Phys. Ed. Grade 15

Actual weights do not deviate markedly from those assigned.

The most significant relation emerging is that between Aptitude for Service and Efficiency Rating as an officer (r equals .34 with values ranging from .51 for Infantry, through .32 for Technical Service to .18 for the Air Forces). While it is clearly recognized that the objectives of the training at West Point are manifold and the data presented are limited by the short period of service as an officer which is covered, and the relatively small number of cases, the need to seriously consider increasing the weight accorded the Aptitude for Service Rating appears evident. Follow-up studies are continuing. If they yield confirmatory evidence, it is not unlikely that they may lead to modification in the weights presently assigned. It has been repeatedly stated that the extremely heavy weight accorded academic standing in determining the General Order of Merit at Graduation (which incidentally determines position on the promotion list for army officers) is a reflection of limited confidence in the Aptitude for Service Rating. To the extent that evidence such as that presented here is accumulated, this attitude may be expected to change. Ultimately, of course, the assignment of relative weights must be somewhat arbitrary since agreement on a single criterion for selecting and evaluating army officers is not now in sight.

ON THE ROLE OF DIRECTIVE AND NON-DIRECTIVE TECHNIQUES IN THE COUNSELING PROCESS

JOHN M. BUTLER
Veterans Administration

It seems evident from the pro and con nature of the discussions in the literature on non-directive and directive¹ counseling that an all-or-none attitude is being taken by some writers with regard to these categories and to the techniques associated with them. Therefore, an attempt will be made in this paper to describe the counseling process in terms of phases in which the application of non-directive and directive techniques may have maximum utility.

In recent years, Professor Carl R. Rogers, his students, and his followers, in their exposition of the counseling process, have tended to define and to delimit this process in terms of social, emotional, and personal problems. At least the issues discussed and the cases presented have revolved around these problems. For them, it appears, the main job of the counselor is to recognize, accept, and reflect the emotionalized attitudes of the client in order that he may accept himself and effectuate a more adequate adjustment by choosing new goals which harmonize with his new-found recognition and acceptance of his spontaneous self. From his understanding of the duties performed by persons known as counselors, guidance counselors, vocational counselors, student personnel workers, etc., who are engaged in aiding individuals to adjust more adequately to their social, cultural, and economic environment, the writer is convinced that such a definition is somewhat restrictive. The majority of counselors known to him are (and should be) concerned with the total adjustment of the individual, whether it is social, emotional, educational, or vocational (economic). If this is acceptable, it follows that in practice the fractionation

¹ The writer does not use the term "directive" in the sense in which it is described in chapter 5, reference 7. His use of the term will become evident later.

of counseling into separate functions such as personal counseling, educational counseling, and vocational counseling is undesirable. This is implicitly recognized even in job descriptions of individuals restricted to "vocational counseling." For example, the booklet, *The Training of Vocational Counselors*, issued by the Bureau of Training of the War Manpower Commission, lists among the functions of the vocational counselor the following:

In a personal interview obtains information about the individual's hobbies, special interests, personality traits, and attitudes. Whenever advisable to do so, also obtains information about financial status, social background, family or personal responsibilities.

Administers appropriate aptitude, interest, personality, trade, and scholastic achievement tests to the individual counselee and prepares the scores for such tests.

It also states that as a basic qualification the vocational counselor must:

possess personal characteristics and skills to interview and maintain continuing relationships in which the cooperation and confidence of the counselee are obtained, and in which pertinent facts and attitudes are discovered and related to a workable plan. Must have the ability to observe personal characteristics in the interview and interpret them in terms of their significance in occupational adjustment.

The purpose of the training unit outlined under the heading, *Personality Adjustments*, is stated as follows:

To give the vocational counselor an understanding of the origin of man's behavior in primitive society and in the influences of environmental and societal factors, the relationship between normal and abnormal behavior with a description of the functions of various agents dealing with abnormal persons. The counselor should be able to identify types of maladjustment and to refer marked personality deviates to the proper specialists.

The implication is clear that vocational counseling is concerned with the personality and with the possible personality disturbances of the counselee as well as with matching men and jobs. Inasmuch as only marked personality deviates are to be referred to specialists it is not illogical to suppose that the vo-

cational counselor is expected to use the data at hand concerning the personality of the client and to help him in the solution of his personal and emotional problems. The conclusion to be drawn, then, is not that the counselor should be a person who seeks only to enable an individual who has personality maladjustments to become more mature and effective, or to enable him to enter a suitable occupation, but one who in addition uses his skills and knowledge directly by facilitating the total adjustment of the individual. This description of what a counselor should do seems to jibe with the duties that many who have been given the title of counselor are expected to perform. Furthermore, it appears to be congruent with what we know about the interrelatedness of adjustment problems. It is so well known as to be a commonplace that counselors are faced every day by individuals who claim to have pressing problems of educational, vocational, or occupational adjustment which are later found to be caused by, or at least closely intertwined with, personality disturbances.

If the counseling process is defined in terms of the total adjustment of the individual, it seems evident that counseling as discussed by Rogers is but a phase in the total counseling process. The purpose of this phase is to facilitate the social and emotional adjustment of the client, without which he cannot face and constructively attack other problems posed by the culture of which he is a living part. This is equivalent to saying that culturally approved behavior will usually result from the application of non-directive counseling and it implies that a relatively constructive attack upon adjustment problems resulting from cultural demands will eventuate. The client's emotional tensions, feelings of aggression, frustration, anxiety, etc., will have been released and will in the process of release have been transformed; he will have gained a new self-orientation and should, with acceptance of the self, be able better to re-orient himself to possible new goals indicated by reports of test results and information with regard to educational or vocational problems.

This phase in which the clients are enabled to accept themselves and which culminates in a constructive attack upon their problems of cultural adjustment will be referred to as the ad-

justive phase and the phase in which clients have shifted to a straight-forward problem-solving approach to the demands of their culture will be referred to as the distributive phase. The latter term is chosen because adjustment to the environment is conceived to be primarily a problem of the distribution of individuals in various socio-economic strata.

It is to be noted that this formulation specifies first that the individual must be ready to use in a positive fashion the information the counselor can give him, including information concerning his interests, abilities, and aptitudes, and second, that this readiness on the part of the client is a function of his personality adjustment. It implies that the interview situation is one in which, if necessary, the client first is enabled to accept himself as a person and recognize and face his personal problems, then receives expert assistance in facing his particular world in terms of competitive cultural and economic demands, and finally learns something concerning the relation of his abilities, aptitudes and interests to these demands.

The distributive phase of the counseling process is one in which the relationship of the counselor to the client increasingly resembles that of the teacher to the student. The counselor may possess test results which will enable the client to assess his position in a scale of ability or abilities, achievement, and interests (as defined by interest tests). By virtue of his experience and training the counselor may provide educational and occupational information which will be of great value to the client who is able to use such information constructively. For example, the university counselor may be confronted by a student who attended a small high school in which the average level of scholastic aptitude was low and in which competition was not keen. The student desires to enter a medical school. His high-school record was good and he had high grades in mathematics, physics, chemistry, and German. However, the *ACE Psychological Examination* indicates that students with a score such as his have less than 3 chances in 10 of graduating from the local medical school. Achievement tests indicate that his achievement in mathematics, physics, chemistry, and German, is very low compared to that of freshman pre-medical students. The *Strong Vocational Interest Blank* shows

A ratings for certain business occupations, high occupational level, and high interest maturity. It also reveals C ratings for psychologist, engineer, physician, dentist, and physicist. On the basis of his limited experience and successful competition thus far, the student believes the practice of medicine to be an attainable objective. The counselor believes that the medical objective is unlikely to be realized because the client's ability and achievement are low compared to those of pre-medical and medical students. Furthermore, his interests as defined by the *Strong Vocational Interest Blank* are more like those of certain types of successful businessmen than they are like those of successful physicians. If this student is well adjusted personally and can react realistically to problems posed by his environment, he should be able to make good use of the information which is presented in impersonal and non-technical terms by the counselor. He may, if his desire to practice medicine is strong enough, take a relatively light pre-medical program, take advantage of "how-to-study" classes, hire special tutors, and apply himself intensively to his work. He may succeed or fail. He may consider other objectives as a result of the information given to him by the counselor, and may discard medical practice as an objective. In any event, he has benefited from the application of "directive" techniques by the counselor because he has learned something about his position with respect to his fellows. He has learned quickly through vicarious experience what must otherwise have been learned over a long period of time, and has been spared the possible traumatic effects of humiliating failure because he has become acquainted with the odds facing him.

It appears that the techniques described by Rogers which seem to consist primarily of defining the interview situation in terms of client responsibility and of accepting and reflecting the attitudes and feelings of the client should, when possible, routinely precede the distributive phase of counseling, which because of its diagnostic aspects and the usual familiarity of the counselor with the fields of occupational and educational information and tests, becomes increasingly "directive" as the phase nears completion. Needless to say, the use of "directive" techniques and the imparting of information to the client *may*

lead to unfortunate results in terms of opportunities lost to him if he is unable to evaluate or to receive information in a positive and constructive fashion. In terms of office procedure this would mean an interview or interviews before the client takes any tests or before searching questions are asked. The counselor must, of course, be able to decide at what point in the process "directive" techniques become appropriate. While this may seem to be a simple matter it makes great demands on his skill, since a primarily diagnostic attitude may dull his sensitivity to subtly expressed attitudes or feelings of the client with a consequent loss of rapport. Such a loss of rapport will result in a distortion of the counseling relationship and the conditions necessary for the effective application of non-directive counseling will no longer obtain.

Clients whose basic problems are emotional in nature often insist that they be given tests and occupational or educational information in order to satisfy needs which are not directly related to occupational or educational adjustment. They say in effect "Here I am, give me your tests; evaluate my personality, ability, and experience; set my goals and give me a push toward those goals." In such cases it is obvious that the administration of tests and the giving of information may often be handled advantageously on a non-directive basis. For example, the available tests or types of tests may be described and the client may elect which he will take. Such a technique should result in less rationalization and defensive behavior on the part of the client when the test results are reported by the counselor. The attitudes of the clients toward the tests and test results are the concern of the counselor; the information derived by the client is only incidental, since it is more than likely to be distorted considerably as a consequence of his needs. To enter the distributive phase of counseling by forcing intellectual give and take on an unready client may result either in a lack of acceptance or in slavish acceptance of the test results, it may allow him to depend upon the counselor for the solution of his problems, and, more than likely, may affect the test-taking process itself.

In the opinion of the writer, more fundamental issues have been overlooked because authoritarian counseling has not been

clearly differentiated from authoritative opinions. This may be a natural consequence of considering only those individuals who have primarily social-personal-emotional problems to be fit subjects for counseling. By defining counseling in terms of the total adjustment of the individual this difficulty can be avoided. For example, Rogers has described directive techniques as follows: (a) The counselor defines the problem in terms of diagnostic or remedial procedures. (b) The counselor indicates the topic to be discussed and leaves its development to the client. (c) The counselor indicates the topic and delimits its development to confirmation, negation, or the supplying of specific items of information. (d) The counselor identifies a problem, a source of difficulty or a condition needing correction, etc., through test interpretation, evaluative remarks, etc. (e) The counselor expresses approval, disapproval, shock or other personal reactions in regard to the client. (f) The counselor explains, discusses, or gives information related to the problem or treatment. (g) The counselor proposes client-activity, directly or through questioning techniques, or in response to the question of what to do. (h) The counselor reassures the client. (i) The counselor interprets the test results but not as indicating a problem, source of difficulty, etc. (j) The counselor influences the making of a decision by marshaling and evaluating evidence and expressing personal opinions. It may be that these techniques are inappropriate when the counselor is confronted by clients with crippling personal problems. However, techniques (a), (c), (e), (g), (h), and (j), are inappropriate in any counseling situation and in the opinion of the writer can only be characterized as bad counseling in any context. To identify such techniques as characteristic of counselors employing directive methods appears to be equivalent to saying that only non-directive techniques are appropriate. This may be true when counseling is restricted to clients with a relatively narrow range of adjustment problems. However, techniques (b), (d), (f), and (i) are not inappropriate in certain situations and, furthermore, there need be no conflict between these techniques and those listed by Rogers as more or less unique to the non-directive counselors, provided a proper sequence is followed in counseling with the client. In the writer's opinion the term

"directive" should have connotations which include such concepts as direction toward a goal, authoritative information, client responsibility, the presentation of possible alternatives, etc. It should not connote counselor directions, suggestion, and "slanted" information based upon the desire of the counselor to have the client move in the direction of a counselor-selected goal.

Summary

The counseling process as defined by Rogers and his followers appears to be almost identical with the concept of psychotherapy. Counseling, however, may be defined in terms of the total adjustment of the individual. Such a definition is inclusive of the concept of psychotherapy and aims at aiding the individual to face and to solve his adjustment problems whatever may be their nature. If the latter type of definition is acceptable, the counseling process is discerned as having two phases which have been called the "adjustive" and the "distributive". In the adjustive phase the emphasis is upon the social, personal and emotional problems of the client; in the distributive phase the focus is upon his vocational, occupational, and educational problems. Non-directive techniques seem to be appropriate to the adjustive phase because constructive impulses leading to the choice of appropriate goals, and hence, better adjustment, are distorted by emotionalized attitudes and feelings of guilt, fear, rejection, etc., which hinder the attainment and even the choice of such goals. Directive techniques, which imply the use of authoritative judgment by the counselor based upon his training and background in the fields of psychology, tests and measurements, occupational information, education, etc., are appropriate to the distributive phase of counseling. The distributive phase of counseling should not be entered until the growth processes inherent in the client result in independent and mature judgments and readiness to choose new goals or to abandon old ones. Non-directive and directive techniques are recommended for use in the adjustive and distributive phases of the counseling process, respectively, because in the former emphasis is placed upon the emotional and social adjustment of the client and in the latter

emphasis is placed upon the intellectual and rational consideration by the client of various alternatives presented by the counselor in addition to the alternatives he may be considering on his own.

REFERENCES

1. Axline, Virginia "How Counseling is Done in the Counseling Center at the University of Chicago" Unpublished paper.
2. Bixler, Ray H. and Bixler, Virginia H. "Clinical Counseling in Vocational Guidance" *Journal of Clinical Psychology*, I (1946), 186-192.
3. Bixler, Ray H. and Bixler, Virginia H. "Test Interpretation in Vocational Counseling," *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 145-146.
4. Curran, A. C. *Personality Factors in Counseling* New York: Grune and Stratton, 1945.
5. Darley, John G. *The Interview in Counseling*. Washington: Retraining and Reemployment Administration, US Department of Labor, 1946.
6. Rogers, Carl R. "The Processes of Therapy" *Journal of Consulting Psychology* IV (1940), 161-165.
7. Rogers, Carl R. *Counseling and Psychotherapy* New York: Houghton-Mifflin, 1942.
8. Rogers, Carl R. "Psychometric Tests and Client Centered Counseling," *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 139-144.
9. Rogers, Carl R. "Counseling," *Review of Educational Research*, XV (1945), 155-163.
10. Snyder, W. U. "An Investigation of the Nature of Non-Directive Psychotherapy." *Journal of General Psychology* XXXIII (1945), 193-223.
11. Rogers, Carl R. "Significant Aspects of Client-Centered Therapy." *The American Psychologist*, I (1946), 415-423.
12. *The Training of Vocational Counselors*. Washington: Bureau of Training, War Manpower Commission.
13. Williamson, E. G. and Darley, J. G. *Student Personnel Work*. New York: McGraw Hill Book Co., 1937.
14. Wrenn, C. Gilbert. "The Interview." *Review of Educational Research*, IX (1939).
15. Wrenn, C. Gilbert, "Counseling With Students" in *Guidance for Educational Institutions*. Part I, 37th Yearbook of the National Society for the Study of Education. Bloomington, Ill.: Public School Publishing Co., 1938
16. Wrenn, C. Gilbert. "Client-Centered Counseling." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENTS*, VI (1946), 439-444.

COUNSELING: ECLECTIC OR SYSTEMATIC?

RAY H. BIXLER

Minnesota Psychiatric Institute

FOLLOWING the appearance of Rogers' *Counseling and Psychotherapy*, numerous articles advocating eclecticism have appeared, largely in protest against the employment of a systematic point of view in interviewing. In attempting to establish the merits of eclectic counseling, these writers have leaned heavily upon certain fundamental premises which are summarized below.

1. The use of the *appropriate* technique at the *appropriate* moment promises more effective results than rigid adherence to any one method.

2. Selection of the *appropriate* therapy from an "armament of therapies" by a skillful worker on the basis of diagnostic and case history techniques gives greater promise of success than systematic counseling.

3. *No one* method is appropriate to the problems of *all* clients.

Although these premises are undeniably sound, it is well to keep in mind the level at which we now deal with the problems of our clients. Diagnosis of emotional, vocational, and educational problems is exceedingly crude. With a relative degree of certainty, diagnoses such as "alcoholic," "vocational indecision," "behavior problem," or "disturbed" can be made. Diagnosis at the next level (anxiety neurosis, schizophrenia, etc.) becomes more difficult. Yet only the most naive counselors would assume that such symptomatic diagnostic labels determine the method of treatment to be employed, little less the technique for any given moment. When an attempt is made to determine causal relationships ("reading problem arising out of parental rejection," etc.), the experienced counselor is frequently taxed beyond the limit of his skills. It should be remembered that even at this level of diagnosis the

counselor has no criteria by which selection of the *appropriate* technique at the *appropriate* moment is feasible. (When etiology can be established, it is often possible, however, to achieve more efficient selection of the appropriate gross method, such as environmental manipulation, versus counseling.)

In the light of our present diagnostic skills, it would seem safe to state that we have no criteria for the selection of techniques at any one point in therapy, nor do we even have criteria by which we can determine which of numerous interview therapies is most effective with any one client.

The use of an "armament of therapies" raises one additional question for those schooled in any systematic approach. The systematist has found that achieving a high degree of skill in the one method he has studied under supervision is quite a task. He is openly apalled at the mere thought of learning to be facile as an analyst at nine o'clock, a client-centered counselor at ten, and an Adlerian at eleven, not to mention the afternoon with Williamson, Horney, Allen, and Hartwell.

The third premise is undoubtedly appropriate. Although the systematist verbalizes the limitations of his method, all too frequently he denies them in practice.

With our present lack of understanding, clinical experience and observation justify certain tentative conclusions:

1. Regardless of the method of counseling employed, success in treatment bears a close relationship to the problem confronting the client. Severe rejection or indulgence pose the most difficult problems of therapy in parent-child relationships, regardless of the therapeutic method. Hysterical personalities usually react favorably to many methods of treatment, while alcoholic and homosexual personalities resist the same therapies. One possible interpretation: It is the client rather than the type of therapy that determines treatment success.

2. There is a lack of evidence to indicate the superiority of any one method.

3. Therapy is largely a non-intellectual process. The resistant child who sits passively through the therapeutic period time after time, the student who is allowed to refuse therapy, and the client who "catharts" get better—all too

frequently for any of our pet theories. The writer now has a passive delinquent (in a very resistant way) who reads a magazine during the interview period. He is coming in under court pressure. When environmental changes were suggested to him, he accepted on the condition that he could continue our contacts! He is making constant improvement. Discussing these experiences with other therapists, the writer finds that they can be multiplied many times.

The writer believes the most appropriate interpretation of this data must place the client's personality and the counselor's attitude above methodology. It appears that what the counselor says pales in significance beside how he says it; that is, how he feels toward the client.

This philosophy is not easily translated into counseling action. Furthermore, the concept has little value in supervised training. One has only to imagine how confused student therapists would be if they were told, "The only thing you have to do is accept your clients." Acceptance must be implemented by techniques which are meaningful to the student therapist and facilitate his comfort with the role of accepting his clients if we are to make constructive strides in imparting our knowledge and understanding to him.

Rogers (3, Chapter III) has attempted to put this process into words. The written word transmits only the intellectual aspects of any technique. Snyder (4) has pointed out that rigid adherence to verbal acceptance without feelings of acceptance results in poor therapeutic gains.

In light of our present understanding, this does not offer any solution to the choice of methods. The eclectic can easily justify his approach by saying:

When we know so little shouldn't we use every technique
• and resource possible?

The nondirectivists can parry with.

Since we know so little and what we know tends to indicate that the key lies in the client, why not give the counselor the security of *one* facile approach which minimizes the danger of traumatizing the client and frees the counselor to be sympathetic and understanding? Or must the counselor through continual selection of this or that method which he hopes is appropriate remain on a cold analytical plane?

Although such outpouring of claims and counterclaims based on feeling rather than evidence might appear ridiculous to the casual observer, it is probably a necessary prelude to progress.¹ By following our own hents under the intense stimulation engendered by strong feeling, we hasten the exploration so necessary to the refinement of therapeutic procedures.

REFERENCES

1. Bordin, Edward S. "Diagnosis in Counseling and Psychotherapy," *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 169-184.
2. Meister, R. K. and Miller, Helen. "The Dynamics of Non-Directive Therapy," *Journal of Clinical Psychology*, II (1946), 59-67.
3. Rogers, Carl R. and Wallen, John L., *Counseling With Returned Servicemen*. New York: McGraw-Hill Book Co., 1946.
4. Snyder, W. U. "'Warmth' in Non-Directive Psychotherapy," *Journal of Abnormal Social Psychology*, XLII (1946), 491-495.

¹ Recent articles by Meister and Miller (2), Bordin (1), and Snyder (4) give promise of progress. Without attempt to evaluate content, the merit of these contributions lies in their concern with integration and refinement of techniques, rather than defense of a method or disturbance because of its deviation from "tried and true" concepts.

AN APPLICATION OF GUTTMAN'S NEW SCALING TECHNIQUES TO AN ATTITUDE QUESTIONNAIRE

KENNETH E. CLARK AND PHILIP H. KRIEDT

University of Minnesota

DR. Louis Guttman and his colleagues have developed a method of scale analysis which claims to permit the measurement of attitudes by the use of a small number of questions, and to provide ready checks to determine whether these questions belong together as different facets of the same issue.¹ Significant features of this new method are, a new set of criteria for determining whether items of an opinion or attitude questionnaire can be considered to constitute a unidimensional scale for which a single total score is meaningful, a technique for re-scoring items so that they will meet the requirements of this criterion more successfully, and, methods for determining a "zero-point," so that the population may be divided more readily into pro's and con's. That Guttman sees the possibilities for wide-spread applications of these methods is indicated by the following quotation: "This approach has wide ramifications not only for attitude and opinion research, but for many other fields like market research, mental testing, and elsewhere where it is desired to quantify qualitative data."²

This report concerns the application of Guttman's techniques to a conventional attitude scale. A twenty-two item scale constructed by Rundquist and Sletto to measure the economic liberalism-conservatism continuum was selected.³ This particular scale was chosen not because it promised to be unidimensional in character, but because it had been subjected to an accepted method of attitude scale refinement, namely, the

¹ Guttman, Louis "The Cornell Technique for Scale and Intensity Analysis" *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VII (1947), 247-280.

² Guttman, Louis, *op cit.*, p. 247

³ Rundquist, Edward Alfred and Sletto, Raymond Franklin *Personality in the Depression: a Study in the Measurement of Attitudes*. Minneapolis University of Minnesota Press, 1936

Likert technique. The present study limits itself to Guttman techniques of content analysis, and neglects the techniques of intensity analysis.

According to Guttman's criterion for unidimensionality, which he calls reproducibility, a scale has perfect reproducibility if from a person's rank on the total scale it is possible to reproduce his response to each item in the scale. Consider a scale of items each of which is scored two, one, or zero, perhaps for responses "Agree," "Undecided," and "Disagree." (The same method can be used for items having any number of response categories.) The reproducibility of any single item is determined by preparing a table in which respondents are ranked according to their total scores, and by recording for each his response to the item in question. If an item has perfect reproducibility, each person scoring two on that item will have a higher total score than any person scoring one or zero, and each person scoring one on that item will have a higher total score than any person scoring zero. Since perfection is rarely attained, the degree of reproducibility is ordinarily calculated.

The prepared table of respondents is scrutinized, and, for each item, two cutting points are arbitrarily selected. When properly placed, these points divide the list of respondents into three sections so as to maximize the percentage of two-responses in the top section, to maximize the percentage of one-responses in the middle section, and to maximize the percentage of zero-responses in the bottom section. The percentage of responses to each item which do fall properly into the three sections set off by these two cutting points is that item's per cent reproducibility. Guttman seems to apply the term reproducibility more to an entire scale, however, than to an individual item. The reproducibility of a scale is the mean of its individual item reproducibilities. A scale is considered to have satisfactory unidimensionality if it has a reproducibility of 90 per cent or more.

The reproducibility of a scale is improved by increasing the reproducibility of individual items. The Guttman method does not improve a scale solely by eliminating poor items and retaining good ones, as do most methods of scale refinement,

but modifies the scoring of items by combining response categories which overlap: that is, response categories which are selected in nearly equal proportions by persons with differing scores on the total scale. Re-scoring would be needed, for instance, if examination of the table in which the respondents are ranked by total score shows that the two-responses for an item all come first, but the one- and zero-responses intermingle. The two-responses should still be scored as two, but the one- and zero-responses should be combined, and all be scored as one's or zero's. After necessary re-scoring has been accomplished for all items, new total scores are computed, respondents are re-ranked, and a new per cent reproducibility figure for the scale is determined.

This technique of scale analysis was applied by the writers to data secured by administering the *Rundquist-Sletto Scale of Economic Conservatism* to 306 students enrolled in a psychology laboratory course at the University of Minnesota. Since the writers wished to determine the extent to which chance factors influence per cent reproducibilities of items, the completed questionnaires were divided randomly into two samples: Group A, those students sitting in even-numbered seats, and Group B, those in odd-numbered seats. Each writer worked independently applying the Guttman methods to one of these two sample groups.

Difficulties were encountered almost immediately upon beginning the analysis of these data. In the first place, the writers found the process of assigning cutting points for computing reproducibilities to be difficult and arbitrary. Guttman's descriptions of methods for computing reproducibilities were found to be inadequate for insuring that each of the writers would use the same criteria for determining these points. While each of several sets of cutting points would give equal per cent reproducibilities, some sets would hold out greater promise than others for improved reproducibility when item responses were combined. In the second place, little agreement was observed between the two samples as to those items which had high reproducibility. Discrepancies were observed in spite of the fact that the distributions of item responses for the two samples were quite similar.

Having observed low and unsatisfactory reproducibilities for the scale for both samples, each writer attacked the problem of combining response categories so as to increase reproducibilities. The necessity for following a standard procedure for the improvement of reproducibility without introducing extraneous factors led the writers to dichotomize all response categories. It was agreed that each would dichotomize by combining 5- and 4-responses and by combining 2- and 1-responses. The 3-responses were to be thrown with the 5,4 or 1,2 group so as to minimize the number of misplacements, thereby maximizing the reproducibility of the item. It should be noted at this point that, had the writers dichotomized their response categories to separate either the 1- or 5-response for one category, and to combine the four remaining responses for the other category, it would have been possible to obtain a reproducibility of 96 per cent for both samples even if all items were completely unrelated to each other. This is true because of the way in which the obtained reproducibility figure depends upon the distribution of responses. As Guttman notes, "... reproducibility of an item can never be less than the largest frequency of its categories. . . ."⁴ Thus, if an item can be dichotomized so that 90 per cent of the responses fall in one category, the item will necessarily have a reproducibility of at least 90 per cent.

Guttman offers no satisfactory solution to this problem of "spurious reproducibility." He says: "An empirical rule for judging the spuriousness of scale reproducibility has been adopted to be the following: no category should have more error in it than non-error."⁵ This seems to be an arbitrary and off-hand attempt to solve what is one of the fundamental weaknesses of the Guttman scaling scheme.

When the results of dichotomization of categories were compared for the two samples, it was found that in only six out of the twenty-two items did the writers agree in the placement of the 3-response. Eleven agreements is the average that would have been expected by chance. Each sample was therefore handled separately, total scores were re-computed,

⁴ Guttman, Louis, *op cit.*, p. 260.

⁵ Guttman, Louis, *op cit.*, p. 261.

individuals re-ranked according to this total score, and reproductibilities re-computed. Original and new item and scale reproductibilities are presented in Table 1. Combining all

TABLE 1
Per cent Reproducibilities for Individual Items, and for the Total Scale, With Five and Two Category Responses, Rundquist-Sletto Economic Conservatism Scale

Item Number	5-Category Responses		2-Category Responses	
	Group A	Group B	Group A	Group B
1	64%	58%	95%	84%
2	49	45	84	75
3	41	44	65	81
4	57	51	70	74
5	61	42	81	81
6	50	65	72	90
7	62	52	81	78
8	52	45	75	71
9	52	45	77	70
10	54	45	78	85
11	44	45	66	73
12	62	55	87	82
13	57	50	72	79
14	59	41	82	78
15	62	58	70	85
16	60	52	89	84
17	59	47	79	84
18	68	55	85	84
19	52	50	81	88
20	58	54	79	83
21	52	45	83	76
22	62	55	76	83
Total Scale . . .	56%	50%	79%	80%
N . . .	149	158	149	158

response categories into dichotomies has increased reproducibility for the scale from 56 per cent to 79 per cent for Group A, and from 50 per cent to 80 per cent for Group B.

Since the lowest possible reproducibility for a scale consisting of five category responses is 20 per cent, and for a scale of dichotomized items, is 50 per cent, it is possible that the observed increase in reproducibility merely reflects the increase in minimum values. (Actual minimum values for Group A were 42 per cent for the 5-category scale and 68 per cent for

the 2-category scale: for Group B these values were 42 per cent and 64 per cent).

These new scale reproducibilities still do not meet Guttman's criterion for satisfactory unidimensionality with either of the two samples. One must conclude, on the basis of the analysis by Guttman's methods, that this scale does not possess unidimensionality, and that, therefore, it should not be used to obtain a single score which would be used to reflect an individual's position on a liberalism-conservatism continuum. In fact, one may interpret Guttman's writings to indicate that the very existence of such a continuum has been denied.

TABLE 2
*Odd-Even Reliabilities of the Rundquist-Sletto Scale Before and After Re-Scoring by Use of Guttman's Methods**

	Reliability	
	Group A	Group B
Original Scale	.87	.87
Scale Modified by Dichotomizing all Response Categories	.85	.88
N	149	158

* The Spearman-Brown Prophecy Formula has been applied.

This conclusion is reached by Guttman's methods in spite of the fact that this scale, whether in its original form, or modified into dichotomous response categories, does possess quite high internal consistency as measured by odd-even reliability. These values are presented in Table 2. These reliabilities approximate the .85 and .82 values reported by Rundquist and Sletto.⁶ Information regarding the internal validity values of individual dichotomized items is presented in Table 3. Item validities were computed using Flanagan's table.⁷ These values are estimates of the product moment correlation coefficient obtained using the upper and lower 27 per cent of the sample. In contrast to the item reproducibilities presented

⁶ Rundquist, Edward Alfred and Sletto, Raymond Franklin. *Personality in the Depression: A Study in the Measurement of Attitudes*. Minneapolis: University of Minnesota Press, 1936.

⁷ Flanagan, John C., "Item Analysis by Test Scoring Machine Graphic Item Counter." *Proceedings of the Educational Research Forum*. New York: International Business Machines Corporation, 1940. pp. 89-94.

in Table 1, these values appear to be very satisfactory. One must conclude either that the Guttman scaling methods do not apply to scales of this type or that Guttman's techniques do uncover weaknesses which are obscured when usual methods of scale analysis are used

TABLE 3

Item Internal Validities Computed by Flanagan's Method For Rundquist-Sletto Scale When Responses are Dichotomized by Use of Guttman's Methods

Item Number	Item Validities	
	Group A	Group B
1	.67	.77
2	.66	.63
3	.40	.42
4	.53	.55
5	.80	.71
6	.51	.80
7	.60	.68
8	.51	.47
9	.74	.47
10	.50	.70
11	.40	.52
12	.54	.74
13	.47	.72
14	.72	.67
15	.57	.82
16	.53	.63
17	.63	.76
18	.46	.48
19	.59	.59
20	.71	.73
21	.69	.60
22	.72	.62
Median Item Validity58	.65
N	149	158

Although the following criticisms are based on a single application of the Guttman technique, the writers feel that several basic flaws in the method are apparent and that an attitude of suspicion towards the general application of the technique is justified. The following specific comments are offered.

1. The concept of reproducibility, and the methods used in computing reproducibility, seem like a throwback to the crudi-

ties of early item analysis methods. Emphasis is placed upon counting; cutting points are not precisely determined. Chance factors play a heavy role in determining the value of the obtained statistic, no estimate of the standard error of a reproducibility value is provided. Although Guttman implies that it is relatively easy to determine how to combine categories, the writers found it quite difficult with the present data, and often disagreed with each other.

2. The writers' purpose in making separate analyses of two different samples was to estimate the stability of the statistic, per cent reproducibility. Guttman states that "about 100 persons will usually constitute an adequate sample of the population to test the hypothesis of scalability."⁸ In the present study two samples of about 150 persons each were used. Little agreement was found between the two samples when item reproducibilities were compared.

It is difficult to appraise the stability of the statistic, per cent reproducibility, since to the writers' knowledge it can be estimated only empirically. Comparison of Flanagan item validities and per cent reproducibility for this study leads the authors to conclude that Guttman's index of internal consistency is no more stable than usual indices of item validity. Flanagan⁹ implies that a sample of 500 is required for reasonable accuracy using his method. It appears that Guttman is not justified in assuming that 100 persons constitute an adequate sample to test his hypothesis of scalability.

3. Guttman's selection of 90 per cent reproducibility as the criterion of scale unidimensionality is not only arbitrary, but extremely misleading. The number of response categories used for scoring each item has a direct effect upon reproducibility. As the number of response categories decreases, the amount of possible error automatically decreases. Thus the Guttman method places a premium on the scoring of responses dichotomously. In addition, scale reproducibility is closely related to the "difficulty," or response distribution, of an item.

⁸ Guttman, Louis. "The Cornell Technique for Scale and Intensity Analysis." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VII (1947), 247-280. p. 249.

⁹ Flanagan, John C. "Item Analysis by Test Scoring Machine Graphic Item Counter." *Proceedings of the Educational Research Forum*. New York. International Business Machine Corporation, 1940. p. 92.

A scale having very high reproducibility may be easily constructed by including only very popular and very unpopular statements. In other words, Guttman's criterion, when used alone, encourages two item characteristics which are not generally considered desirable.

4. Guttman's method for estimating reproducibility is so designed as to maximize the statistic obtained, capitalizing on chance error. It seems to the writers that a measure of internal consistency should be neither an over- nor under-estimate, but rather a "best estimate" of the relationship between an item and the total scale.

5. When Guttman talks of reproducibility, he is talking, apparently, of internal consistency. Although he has discussed some of the theoretical implications of reliability in a recent article¹⁰ he has not yet presented any practical analysis or empirical evidence of the ways in which his methods supplement, replace, or otherwise relate to the usual methods for scale analysis. In this study, little relationship was observed between Guttman's measure of the relation of an item to a total score, and Flanagan's index of correlation between an item and the total score. Also, application of Guttman's scaling method had no apparent effect on odd-even reliability.

Guttman's methods of scale analysis may prove to be of value in the fields of opinion and attitude research. It would appear, however, that these methods are still very crude and are not yet ready for widespread use. The present study indicates that in areas of general attitude measurement, at least, serious limitations may exist. Further research is needed to determine the circumstances, if any, under which the use of these methods is warranted.

¹⁰ Guttman, Louis. "On Festinger's Evaluation of Scale Analysis." *Psychological Bulletin*, XLIV (1947), 451-465.

COMPARISONS OF SELF-RATING, PEER-RATINGS, AND EXPERT'S-RATINGS OF PERSONALITY ADJUSTMENT

MARGARET G. POWELL

Children's Services

A CLIENT'S complaints usually form the starting point for counseling, and his feelings of adjustment are frequently used as a measure of progress during counseling. Therefore, it is important to determine the validity of his statements. Are they influenced by a lack of self-insight or distorted by defense mechanisms? Inaccurate or incorrect clues given by the client may result in a loss of valuable time and effort, and may possibly end in an incorrect diagnosis. It seems important to evaluate a person's insight into his own personality adjustment by checking against other measures.

One such measure is the counselor's observations. But he does not have extensive opportunities to observe overt symptoms of maladjustment since he sees the client only in an interview situation. Symptoms of personality maladjustment are usually too closely interwoven into the person's daily experiences and his entire social milieu to be adequately observed in one, two, or three sessions in a limited counseling situation. The counselor usually, therefore, questions the client in an attempt to gain information to supplement his own limited observations. However, the counselor may feel that the client's statements and his own observations are also inadequate and so seeks the observations of additional judges, i.e., the client's associates, co-workers, teachers, friends, or employer.

What is the relationship between information gained from these different sources? Is the picture a client may have of himself the same as that held by the counselor or other expert judge? Furthermore, how much agreement is there between his conception of his own personality and the opinion about him which is held by his friends?

It is a common practice to use only one or two measures of adjustment in making a diagnosis. Perhaps most frequently the counselor uses his own observations and the self-rating of a client to make a diagnosis. Similarly, research into the outcome of therapy is commonly determined by the use of one or two measures. A case is considered as successful following therapy because the client considers himself as better adjusted, or because the counselor considers him as better adjusted, but rarely are additional measures used.

The purposes of this study are, therefore: (1) to determine the relationship between self-insight into adjustment and the "real" nature of the person as shown from ratings of peers and of an expert and (2) to determine the relationship between three different approaches to the measurement of adjustment, i.e., a self-rating, a peer-rating, and an expert's rating.

The field of personality testing has attracted many investigations. Ellis¹ has reviewed many of these studies which attempt to validate self-rating questionnaires. Although some of these studies were similar in purpose to the present study, they investigate only the amount of self-insight individuals have into their adjustment as compared to ratings by other judges. Ratings by more than one class or type of judge were not included in these studies. Either the problem was not considered, or else it was assumed that a person's adjustment as rated by a judge seeing the subject in one light would be the same as that by a judge seeing the subject under other conditions. Such an assumption does not seem to be justified by ordinary observations. The purpose of this study, then, was to investigate the relationship between three measures of adjustment, i.e., rating by self, by the subject's peers, and by their dormitory counselors.

Methods of Investigation

This study was made in a large dormitory which housed approximately five hundred and sixty girls at Ohio State University in the Spring of 1946. For purposes of organization

¹ Ellis, Albert. "The Validity of Personality Inventories." *Psychological Bulletin*, XLIII (1946), 385-440.

these girls were arranged in corridor groups determined by the location of their rooms on the different floors.

Since the large number of girls in the building made it impractical to include all of them in this study, eight corridors totaling one hundred and forty girls were selected as a sampling. These eight groups were selected on the basis of three criteria: (1) The corridors should be large and of the same size in order to give more reliable and comparable peer-rating scores. (2) Each expert, i.e., the four dormitory counselors, included in the study should have only two corridors to rate, so that each should rate a small enough number for all to be carefully done. (3) The corridors selected should have none, or but few, girls who were new since the beginning of the school year so that the peer and expert ratings should not be influenced by a lack of knowledge about any girl.

Testing Materials.—In order to gain a self-rating score or an indication of how well adjusted or poorly adjusted the girls believed themselves to be, the *Bernreuter Personality Inventory* was used. This inventory is representative of the self-rating devices used in counseling practice. While the test does not directly ask for self-evaluation in terms of others, the subject usually answers in terms of his feelings in comparison to what he thinks others feel, and the test percentile, which is a group comparison, is frequently used as a measure of the degree of adjustment.

Two scales of the *Bernreuter Personality Inventory* were used: The B1-N, a measure of neurotic tendency; and the F2-S, a measure of sociability. These two scales measure two aspects of self-feeling and, because of the substantial correlations which have been found to exist between them and the other Bernreuter scales² these two scales are representative of the test as a whole.

On the staff of the dormitory were four Assistant Head Residents who were chosen as the experts for purposes of this study. These Assistant Head Residents cared for the girls on their corridors during illness, supervising the girls in all of the regulations of the hall, counseling and guiding them in

² Bernreuter, Robert G. *Manual for the Personality Inventory* Stanford University. Stanford University Press, 1935

whatever way appeared advisable, and establishing a friendly relationship which would encourage their confidence. This study was made during the Spring Quarter of the school year so that these counselors had been responsible for the girls on their corridors for several months.

The "expert's rating" was obtained by asking the counselor in charge of a corridor to rate the degree of adjustment for each of her girls on a rating scale. These experts were instructed in the use of the scale and warned of the dangers of the "halo" effect.

The expert's rating scale was constructed as follows: On the basis of a thorough review of the literature, items were selected which would sample all of the major aspects of adjustment and differentiate between general adjustment and maladjustment. The final scale was composed of seven traits on which each girl was to be rated. Opposite each item, a rating continuum was set up so that the counselor could mark an "X" on the graphic scale to describe the degree of the girl's behavior on this trait.

These rating scales were scored by securing a numerical score for each of the seven traits rated according to the dot on which the "X" was marked, each dot corresponding to one point. The score on each trait might range from 0 to 40. The sum of these seven scores became the score recorded for each girl.

To establish the reliability of this scale, three other counselors were each asked to rate one of their corridors of girls on the scale. After a period of four weeks they again rated the same girls on this scale. Comparisons of the two ratings showed a median reliability coefficient of .76 for the three groups computed by the rank difference method.

Through day-by-day contacts the girls on the same corridor became well acquainted with each other and it appeared evident that these girls living, working, and spending leisure time together should be capable of judging the personality adjustments of one another.

Two different measuring instruments were used to obtain peer-rating: a "Guess Who" test to measure general adjustment and a sociometry test to measure popularity. The

"Guess Who" test was constructed by re-stating the items included in the expert's rating scale. Of the twenty items included on this test, ten are associated with a well-adjusted person and the other ten with a poorly-adjusted person. On this test the girls were to consider their corridors as a whole and were to insert in each blank the name of the person on their corridors whom they thought most nearly fitted the description. If they could think of no one who fitted the description, the item was to be left blank.

The "Guess Who" test was scored by giving one point to a girl each time her name was inserted for a trait associated with adjustment, and a minus one for each time she was mentioned as having a trait which was associated with maladjustment. The algebraic sum was then computed for each girl, and used as the score on the test.

The Sociometric test was composed of six activities for which the girls were to name from among the other girls on their corridor a first and second choice for companion in various activities. They were also asked to name the girls whom they would reject as a companion in these same activities, with space given for a first and second choice.

The Sociometry was scored by giving weightings to first choices over second choices and the total score was the algebraic sum of the positive and negative weighted votes.

Analysis of Results

The frequency distributions for each corridor on each of the tests were inspected. First, it was obvious that the subjects were homogeneous in the characteristics of sex, occupation, and age. Second, practically all of the frequency distributions took the form of a wide scattering, so the ranking of scores in order could be done with some assurance that unreliability of test scores would not greatly influence these rankings. The distribution of the F2-S scores indicated a clustering at one extreme which increases the unreliability of this measure for comparison with other tests.

Rank difference correlations were computed between each of the measures on each of the eight corridors. Thus, eight rho's were obtained for each test comparison. The median of these

correlations is taken here as the measure of the degree of relationship between the different measures of adjustment.

Validity of Self-Insight. An analysis of the student's rating of her "neurotic tendency" with the expert's rating of her adjustment and with the peer's ratings of adjustment and acceptability produced only very low positive correlations. The median correlations between the B1-N and other tests were: B1-N-Expert, .16; B1-N-"Guess Who", .24; B1-N-Sociometry, .14.

In comparing the F2-S scale of the Bernreuter to the other measures of adjustment, similar results are found. Practically no correlation was found between the expert's rating and the F2-S, the median being .04. The median between the F2-S and the "Guess Who" was .05, and between the F2-S and Sociometry test was .01. Since the F2-S rankings appear less reliable than the others because of a clustering of low scores, relationships between it and the other measures might be slightly higher with a wider scattering of scores. The low relationship between this scale of self-rating and other measures strengthens the evidence of lack of self-insight as noted in the intercorrelations of B1-N and other measures.

Other possible causes of these low correlations should be considered. First of all the low correlations may be due in part to the unreliability of the testing measures. However, the reliability of the B1-N scale has been reported as .91 and .88, and for the F2-S scale .78.³ Studies previously made which have used similar "Guess Who" and Sociometry tests have generally found reliability coefficients of .90 and above.⁴ The reliability coefficient of the Expert's Rating Scale established by test re-test method was .76. These coefficients indicate that the tests used in this study are quite reliable, so that the low relationships existing between them would not appear to result very much from unreliability of the measures.

Another explanation of the low relationships found between these measures may be that in spite of attempts to make them comparable, the tests may be measuring different aspects of

³ *Ibid.*

⁴ Tyron, Caroline McCann. "Evaluations of Adolescent Personality by Adolescents." *Monographs of the Society For Research in Child Development*, IV (1939), 1-83.

adjustment. At our present stage in the development of personality testing the validity of each of these measures cannot be positively established. However, they are used in general practice today as measures of adjustment as used in this study. Lack of self-insight is another possible reason for the low correlations between the rating measures, and appears to be the main reason for the results which were found.

Relationship Among Different Measures of Adjustment

How many measures are necessary for an adequate measurement of adjustment? Do two different tests filled out from the same source, i.e., two peer ratings, give very similar results? Do ratings by different sources agree?

Correlations between the Sociometry test and the "Guess Who" test (both peer ratings) were high and positive. They ranged from .50 to .98 with a median correlation of .86. These results indicate that there was a higher relationship between peer ratings of acceptability and reputation of most of the corridors.

Two other measures of adjustment from the same source are the B1-N and F2-S scales of the Bernreuter, both measures of self-rating. Correlations between these measures ranged from a high of .58 to a low of $-.04$. The median correlation of .40 indicates a rather low relationship between the two measures. These two scales were chosen for this study because they were said to measure somewhat different aspects of adjustment. Therefore, although they both use a self-rating approach, the tests differentiate fairly well between two different personality characteristics.

How well do the three different sources of judgement agree? The median of correlations between the Expert's Rating Scale and the Sociometry test was .34. The median correlation between the Expert's Rating Scale and the "Guess Who" test was .40. These measures show that, in general, measures of adjustment obtained from judges seeing people in a different light show some, but rather low, agreement. And it will be recalled that the median correlations between self-insight and the ratings of these outside judges were very low. It is concluded, then, that all three approaches to the measurement of

adjustment, i.e., self-rating, peer-rating, and expert-rating, are needed in any attempt to measure adjustment adequately.

Comparisons of Extreme Cases

Although there is little relationship between these approaches in judging the degree of adjustment over the entire distribution, there is the possibility that extreme cases might be more consistently located. That is, a girl who was rated as outstanding in adjustment or who was well liked, should, if anyone, feel well adjusted. In order to investigate this possibility, the worst and best-adjusted girls on each measure in every corridor were located and their scores on the other tests studied.

A great deal of disagreement was found between the median ranks of the best and worst girls on each criterion test as compared with their median ranks on other measures. This is shown by Tables 1 and 2. For example, the median rank of the girls who rate themselves as worst on the B1-N would be 18, but their median ranks were 6, 7, and 11 on the expert's ratings, the "Guess Who," and the Sociometry, respectively. Median ranks of 6 and 7 indicate a slight tendency for the expert and the peers to rank these girls as better than average, and the rating of 11 on the Sociometry is not far from the mid-point of 9.5.

Similarly, the median ranks of girls who considered themselves as best adjusted on the B1-N showed no agreement with the median ranks of these girls on other measures. Their median rank would be 1 on the B1-N, but was 7.5 on F2-S, 8.5 on Sociometry, 9 on the "Guess Who", and 9.5 on the expert's rating. The same lack of similarity is found in comparing the ranks of the best and worst adjusted on each criterion test with their ranks on other measures as is seen in Tables 1 and 2.

The evidence from these comparisons indicates: (1) that there is little agreement in extreme cases between self-insight into adjustment and the ratings of others; (2) there is little agreement as to the most maladjusted or the best-adjusted girls on each corridor as rated by self, peers, or an expert. Not only do these different approaches to adjustment show little relationship in the entire distribution, but also they fail to agree in locating the extreme cases of good adjustment or maladjustment.

These results are somewhat startling when we consider that extremely abnormal cases are sometimes located from a group by the use of only one approach to the measurement of adjustment. Our evidence indicates that such an individual may be anywhere within the distribution if rated by a different approach. While such a lack of agreement in locating extreme cases would not be expected from measures which all purport to measure the same thing, the present evidence indicates the need (1) for helping individuals gain better self-insight and

TABLE 1
Median Scores on Other Tests of Worst Adjusted (rank of 18) on Criterion Test

Test	B1-N	F2-S	Sociometry	"Guess Who"	Expert
B1-N.....		17	11	7	6
F2-S.....	12		11	7	9
Soc.....	10	10.5		15.5	13
"GW".....	13.5	12.5	16.5		14.5
Ex.....	15	11.5	15	14.8	

TABLE 2
Median Scores on Other Tests of Best Adjusted (rank of 1) on Criterion Test

Test	B1-N	F2-S	Sociometry	"Guess Who"	Expert
B1-N.....		7.5	8.5	9	9.5
F2-S.....	5		8.5	6	7
Soc.....	10	13.3		1.5	5
"GW".....	6.8	11	1.5		6.5
Ex.....	6.8	6.8	4.5	3	

(2) for using more than one approach in judging the degree of adjustment either before or after counseling.

The results of this study suggest that in counseling it is important to remember that self-insight may have very slight relationship to a person's adjustment as determined by other measures. If other measures are considered, the final diagnosis made by the psychologist may be in direct disagreement with the complaints expressed by the client. On the basis of self-diagnosis it is possible for the wrong persons to be selected for therapy. Also on the basis of self-diagnosis, those who are in need of treatment are often not identified. Thus it seems that self-insight cannot be relied upon as the sole method of selection for therapy.

Summary and Conclusions

The results of this study showed:

(1) There is a lack of self-insight into one's own adjustment as determined by comparisons of self-ratings to ratings by peers and experts; median intercorrelations on eight corridors, between measures were from .01 to .24.

(2) Intercorrelations showed there is little agreement between any two of the measures, except between the "Guess Who" test and the Sociometry where the median of correlations on the corridors was .86.

(3) There seemed to be more consistency in ratings from one source of judgment on two rating devices than between different sources of judgment. For example, the high correlation of .86 between the two measures of peer-rating and the correlation of .40 between the B1-N and F2-S.

(4) These three approaches to rating adjustment not only disagree over the entire distribution of cases, but also disagree in locating those girls who are extremely maladjusted or very well adjusted.

Since the three different measures do not agree in locating cases of maladjustment, no one measure should be taken as adequate evidence on which to base a diagnosis, or on which to determine the outcome of therapy which has been undertaken. It appears that three or more different sources are generally needed to give an adequate picture of one's adjustment.

A REVIEW OF SOME PROBLEMS RELATED TO THE MAIL QUESTIONNAIRE TECHNIQUE

RALPH D. NORMAN
Princeton University

IN all probability, more studies have been made with the questionnaire than with any other instrument in the field of social investigation. Wide application of this research tool occurs in sociology, education, psychology, economics, politics, public opinion measurement, and other areas of social science. This article will attempt to survey some of the more recent literature concerning only the mail questionnaire, discussing problems peculiar to this instrument. Fundamental questions (reliability, validity, wording of questions, etc.) anent the questionnaire method in general are too numerous to cover here, since not only are there many different types of questionnaires, but the field has been covered capably elsewhere. There are good discussions by Symonds (31) and Blankenship (3) and in the periodic reviews which have appeared from time to time (11). Basic considerations of methodology are also available in Kelley (12) and Koos (13). Recently McNemar (16), Crespi (6), and Conrad (5) have debated the questionnaire technique as applied to opinion-attitude research, and Ellis (8) has reviewed the validity of personality questionnaires.

I. *The Problem of Returns in Mail Questionnaires*

To be considered here is the problem of returns, independent of the characteristics of respondents vs. non-respondents which will be considered in section two of this paper.

Percentage of Initial Return.—The percentage of return is influenced by a great many factors, among which may be mentioned the use of follow-ups, inducements or appeals of various sorts, the time when the questionnaire is sent, etc. Reviewing only a sample of the various investigations which have reported returns on a *single* questionnaire or the *first wave* of several

waves of questionnaires we have percentages of returns reported by various workers as follows:

Lazarsfeld (14), 15%; Ram (22), 15.25%; Suchman and McCandless (30), 17%; Shuttleworth (26), 19%; Rollins (25), 23%; Lindsay (15), 27%; Stanton (29), 28%; Ball (1), 40%; Reid (23), 42%; Mudrow (18), 60%; Toops (34), 65%; Kelley (12), 76%.

This represents a wide variation indeed, some of the causes of which are suggested in the discussion which follows in the remainder of this paper.

Follow-Ups.—The percentage of return is affected markedly by follow-up or other motivational devices. Using follow-up letters only, Toops (33) achieved on one occasion almost a perfect set of returns, or 93%, and on another occasion (34) a perfect set or 100% return. In the latter instance, by means of six follow-up letters, he secured the following cumulative percentages of response after each follow-up letter: 65, 83, 89, 97, 99, 100. Reid (23) brought his percentage up from 42 to 69 by means of three successive waves of questionnaires. Stanton (29) also used a follow-up in his study, as did Suchman and McCandless (30). The latter used three waves, the last of which was a telephone wave; the results were 17%, 34%, and 97% (on the telephone wave). In a second study by the same writers, an original mail questionnaire followed by three waves of follow-ups brought returns as follows: 44%, 46%, 50%, 67%, all together returning 95% of the total. Although Sletto (28) found that post cards were as effective as letters in making follow-ups, Lindsay (15) claims that a follow-up card and letter are more effective in their joint effect of pulling returns than the initial letter and questionnaire. He suggests that later follow-ups come at greater and greater intervals of time. By this procedure in a later questionnaire, he received, with three follow-ups, 91.5% or returns against an initial reply of 27% on a questionnaire requiring a half-day to fill out. Patterson (21) in his review of work on the questionnaire, advocates a vigorous follow-up to increase returns, and Toops (35) encourages an intensive follow-up technique which is designed to touch motives as viewed from the angle of the recipients. Concerning other motivational devices, Weaver (37) found that the inclusion of

a duplicate copy of the questionnaire increases the percentage of return. In another instance the same investigator tripled his returns by claiming that no reply was desired!

Other factors, different from the characteristics peculiar to the respondent himself, which may influence the amount of returns may be enumerated as follows:

Reward. Hancock (19) used four methods of obtaining responses to a questionnaire dealing with attitudes toward retail stores. These were (a) a mailed questionnaire with no reward for filling it out and returning; (b) a mailed questionnaire including 25 cents to be kept by the recipient; (c) a mailed questionnaire with the promise that 25 cents would be sent to the respondent if he returned the questionnaire filled out; and (d) a personal interview questionnaire. The percentages of return for each of the methods above were, respectively, 10%, 47%, 18%, and 86%. He concludes that the second method is the best in terms of unit cost and accuracy of results. Shuttleworth (26) found that the enclosures of 25 cents in a simple questionnaire sent out by mail brought 51.6% of replies while the same questionnaire without the coin returned only 19.1% of replies. In the latter case, only a personal letter and a stamped envelope were used. Rollins (25), however, found that when he used an attached pencil with the suggestion that the recipient keep it, he received 23% return on the first wave and only 13% on the second. In this case perhaps the effect of enclosing the pencil on both occasions negated the effect of the "bribe." Care must be exerted, however, in using rewards which may be construed rather as bribes. One is not sure how much offense would be taken at this technique by an intelligent group of respondents.

Form of the Questionnaire.—Moore (17) received an average return of 61% with typewritten questionnaires and 51% with duplicated ones. Toops (36) contends that proper lay-out, capitalization, and paragraphing improves the readability of the questionnaire. Blankenship (3), too, contends that general appearance will be one of the "determining factors" in inducing returns. Neither offers any research to substantiate these claims. Symonds (31) asserts that the design of the questionnaire is most important and urges the use of a single sheet rather

than two if there is enough room. Attention must be paid, he says, to future statistical treatment by machine methods, especially if the sample is large and there are many questions used.

Knowledge Concerning the Sender Toops (34) states that recipients who are best known to the sender reply earlier than strangers. Kelley (12) maintains that there is a greater likelihood of not getting a useful answer if the recipient is not well known to the sender. In this connection, Boring's (4) injunction concerning questionnaire-answering urges this as one of the points to be observed in considering whether or not to reply. The National Education Association (19), surveying percentages of reply received by various classes of questionnaire circulators, reports the following interesting table:

<i>Originator of Questionnaire</i>	<i>Median % of Reply</i>
City supt. or member of central office	75.4
State office.	71.0
School principal or other school employee	66.0
Publishing firm.	61.0
College professors or students	60.4
Private association or foundation	58.0
Miscellaneous.	58.5

Thus, there would seem to be some relationship between the originator and the percentage of reply on a rough prestige hierarchy.

Time of Issuance of the Questionnaire. Early in the week and early in the school year seem to be the best times for receiving replies. Toops (33) reports that the highest returns in his investigation were received on Monday and Tuesday and only about half as many on each of the other four days of the week. The National Education Association (19), in the study cited above, presents a table which indicates that an average of 75% return is to be expected if the questionnaire is issued in September, October, November, or December, but only 64% average return will accrue in January, February, March, April or May. This, of course, applies to school questionnaires. The above results appear to suggest strongly that in dealing with potential respondents it is better for the investigator to make some preliminary inquiry concerning whether or not his

respondents would be expected to receive the questionnaire at a time which is convenient for reply.

Length of the Questionnaire. - That the length of the questionnaire may have something to do with the percentage of return may be presumed on an a priori basis. Recipients do not wish to be annoyed with a long questionnaire. Toops (35), e.g., cautions that undue length is a symptom of "slovenly technique." There is some experimental evidence for this contention as noted in the following table presented by the National Education Association (19) in their summary of questionnaire-answering behavior:

<i>Number of Items of Information Requested</i>	<i>Median % of Reply</i>
5 or less	78.5
6-10	71.0
11-15	71.0
16-20	68.5
21-30	66.0
31-40	61.0
41-50	71.0
51-75	76.0
75-100	66.0
101-500	54.3
501-1000	
1800-1900	

Thus, except for a reversal in the middle of the table, it would appear that there is a rough relationship between the length of the questionnaire and the amount of reply.

2. Respondents vs. Non-Respondents to Mail Questionnaires

Most cautious investigators who have examined questionnaires, especially the mail type, have realized that the matter of who returns a questionnaire is a vital one to any investigation which undertakes to use the postal inquiry as its basic instrument. The problem is: Are those individuals who reply to a questionnaire representative of the population to whom the questionnaire was sent? A corollary of this question is one other: Is there any difference between early and late respondents? Also related to this problem is that of the means to secure increased, or more representative returns, discussed on the preceding pages.

Thorndike in his critique of the questionnaire method says:

Psychological questionnaires are commonly answered by only a limited number of those who receive them—namely, by the individuals to whom the questions especially appeal and who have something to report. . . . The replies thus represent an extremely partial sampling of people in general. (32, p. 33)

Symonds, too, raises this question, commenting upon it in greater detail:

It should be remembered that in most inquiries the choice of those to whom schedules are sent is a delicate issue. When a considerable number fail to be returned, the representativeness of the sampling may be seriously deranged. The investigator must always ask himself if there is a probability that those who failed to answer did. If those who do not answer fail to do so because of lack of interest or because of ignorance, there is a strong probability that their returns, if they could be obtained, would alter the results. Usually when returns are received from only a small percentage of those to whom question blanks are sent, some estimate must be made which allows for this factor of selection. (31, p. 129)

Research is almost unanimous in corroborating these comments by Symonds concerning the differences between those who do and those who do not reply to questionnaires. The only dissent comes from Toops (34) who states that the essential conclusion of his study was not materially changed through the replies obtained by follow-up letters, although he does say that with respect to other conclusions this would not necessarily hold. In an earlier study, Toops (33) found that neither the early nor the late returns were a biased sampling of all returns, so far as the most essential question of the study was concerned. On the other hand, there are several studies which do not bear out Toops' contention. Ball (1), using follow-up letters or questionnaires containing six simple questions which were sent to the parents or guardians of delinquent boys released from a boys' school in San Francisco in 1926, 1927, and 1928, reports that parents are more likely to answer the questionnaire if their report is favorable. Kelley (12), in sending a questionnaire to faculty members of a university, noted that he was not likely to get a useful answer if the recipient held a minor position of one-year tenure. More elaborate studies by

other investigators are striking in their unanimity concerning the differences between respondents and non-respondents.

Stanton (29) for example, using a three-page mail questionnaire which was sent to a representative list of 11,169 United States school teachers, inquiring among other things about their possession and use of classroom radio receiving facilities, found that the replies of the teachers who needed a follow-up differed markedly from those who made an original reply. There were significantly less who owned a radio in the classroom in the former group. Reid (23) in another questionnaire on the use of radios in schools, this time in Ohio, followed up a sample of non-respondents. There was a consistent decrease in the percentage of affirmative responses with each succeeding group of respondents in three successive waves. He felt that probably the primary reason for this was lack of interest; those principals interested in radio tended to answer and to return questionnaires on radio more readily. Similar results on an occupational survey were found by Shuttleworth (27). He sent questionnaires to 327 technology and chemistry majors graduated from the City College of New York in 1936. Repeated efforts to contact these alumni brought information concerning all but four of the 309 for whom there were sufficient addresses. The first 184 returns showed only .5% unemployed, while the next 121 returns showed 5.8% unemployed. In the case of the technology majors, 65 early returns showed only 15.4% employed outside their field of training; 54 later returns, however, showed 29.6% thus employed. A similar trend also held for chemistry majors. Former college students were also followed up in another study by Pace (20). He found that non-returners (of questionnaires) for both sexes were heavily weighted with students leaving at any time prior to the completion of nine quarters of work. Graduation from the university was also related to return versus non-return. He investigated further the late returners versus the early returners and found that late versus early return was somewhat prognostic of return versus non-return. He states that whether or not a person will return a questionnaire and when he will return it depend on a favorable combination of all of the factors which influence questionnaire returns: interest, conscientiousness, promptness, time available,

pleasurable associations with the source of the questionnaire, sufficient lack of embarrassment with one's present status to be willing to report that status, and many other factors. Reuss (24) also reports similar results. A follow-up study of 1,097 college students was made by means of a mailed questionnaire. Marked differences existed between respondents and non-respondents; higher intelligence scores and scholarship, loyalty or ties to the questionnaire sponsor, and a rural background were associated with a positive tendency to respond.

Suchman and McCandless (30) find substantially the same pattern in another type of research, that concerned with radio-listening behavior. In two separate investigations reported in the same article, they studied (a) a random list of 600 women to find out whether they listened or didn't listen to a child-study broadcast program, and (b) a random list of 700 listeners who subscribed to a classical music program bulletin. In both cases differences between the two groups who replied or did not reply to a mailed questionnaire concerning the programs were ascribed by the authors to interest or familiarity with the topic under investigation and to the education of the respondent. Rollins (25), seeking to find out for his magazine how many readers had ever flown on a commercial airline, found that 17% among the first wave of respondents had flown, but only 7% among the second. He concluded that it was obvious that those who were more interested in flying were likely to answer the first time. Toops (35) gave early support for this matter of interest, for, in his examination of 135 questionnaires studies, he offers advice to select for study a topic in which the recipients are "vitally interested" in order to assure a high percentage of returns.

Franzen and Lazarsfeld (9), writing on the mail questionnaire as a research problem, undertook to find out what caused people to balk at answering mailed questionnaires. They sent out questionnaires to 3,000 representative subscribers to *Time* magazine, 1,000 each of three slightly different forms; 505 persons who replied and 882 who didn't reply were personally interviewed. Their results indicate that mail questionnaires can produce valid samples of comparatively homogeneous

groups and that the answers to some questions given in a mail questionnaire are more informative and more freely given than the answers the same people give to the same questions when face to face with an interviewer. There is a freer admission, e.g., by mail, of unusual activities or interests, of buying power, of personal income, than is made to the personal interviewer. This study is unique in showing, in this fashion, that greater validity *may* sometimes be achieved by the mail questionnaire - it is of great value to consumer researchers, for example, to know that the impersonal mail questionnaire will be more fruitful on these items. On the negative side, these authors indicate that statistically significant biases do occur in mail returns, and the possibility that they may be large enough to prejudice generalization must be provided for in any survey.

In a somewhat different fashion, a German investigator, Mudrow (18), also proceeded to determine whether there were characteristic differences between the group who answered questionnaires and those who did not. Proceeding by means of personal contact, letters, etc., he studied the results available from 4,000 answered questionnaires, the number returned out of 10,000. The following reasons, the author claims, are the more important for not answering: indifference, negligence, suspicion, lack of understanding, fear of affecting their status (given by women), and minimizing the importance of personal data. Opposed to these characteristics are those of the group who gave replies. Considering the problem from the constitutional-biological point of view, the author suggests the possibility that the non-answering group might have a larger percentage of cyclothymic personalities.

Excellent illustrations of bias in questionnaire surveys in the field of public opinion have been presented by Benson (2). He points out (a) that under certain conditions people in the higher income and educational brackets will return mail ballots with consistently greater frequency, (b) that a disproportionately greater percentage of people who were strongly opposed to the 1937 Supreme Court "packing" proposal by President Roosevelt made a special point of registering their protest in a mail survey, and (c) that the mail ballots used by the *Literary Digest*

in polling the 1936 Presidential election erred by the amazingly high amount of about 20% because of strong prejudice on the part of the respondents.

A recent article by Edgerton, Britt, and Norman (7) presents striking objective evidence of bias in mail returns. They conducted a follow-up study of all male contestants in the First Annual Science Talent Search by means of a mailed questionnaire. There were three classes of contestants which emerged as a result of the search, which was a nation-wide contest to discover scientific talent. These three classes were winners, honorable mentions, and others, or "also rans." These investigators found that winner contestants make almost perfect returns of the questionnaires for each of three follow-up years, honorable mention contestants make the next largest percentage of returns, and the "also rans" have the lowest percentage of returns. A second part of their study concerning the *Science Aptitude Examination* scores of the "also-rans" for four follow-up years reveals that if individuals continue to reply faithfully to the questionnaire year after year for four years they tend to be superior on this test to those who make no reply or who reply in any lesser degree of succession. Moreover, there is a rising mean value on the aptitude test in direct relationship to the frequency of replies, disregarding succession. Using the fourth year of follow-up only they found that early respondents had a higher, although statistically not significant, mean score than late respondents.

The implications of these findings concerning the problem of who answers questionnaires are obvious. Since ties to the questionnaire sponsor or interest in the subject under investigation are productive of higher returns, intensive and vigorous follow-up is a basic tenet of all mail questionnaire research. Otherwise, the tendency will be to obtain replies from those who have a special interest in the subject under study, or who exhibit some characteristics different from the non-respondents or from the casual or indifferent respondents.

Summary

In reviewing the literature concerning some of the problems peculiar to the mail questionnaire technique, it may be con-

cluded that the percentage of return on a single questionnaire may vary widely. This percentage is decidedly affected by various devices and by the characteristics of the respondents. Follow-ups in general increase the percentage of return, as do simple rewards. There is little research evidence on the form of the questionnaire as related to returns but, as regards knowledge concerning the sender, there is some evidence for relationship between the originator and percentage of reply on a rough prestige hierarchy. Time of issuance of the questionnaire, especially in education, is of import. Undue length is to be avoided since the amount of reply may be decreased by a large number of questions. Those who respond to a mail questionnaire have been found, almost universally, to differ radically from those who do not reply. It is probable that late respondents also differ from early respondents.

REFERENCES

1. Ball, R. J. "The Correspondence Method in Follow-up Studies of Delinquent Boys." *Journal of Juvenile Research*, XIV (1930), 107-113.
2. Benson, L. E. "Mail Surveys can be Valuable." *Public Opinion Quarterly*, X (1946), 234-241.
3. Blankenship, A. B. *Consumer and Opinion Research*. New York: Harper Bros., 1943.
4. Boring, E. L. "When and How to Reply to a Questionary." *American Journal of Psychology*, XXXVII (1926), 632-633.
5. Conrad, H. S. "Some Principles of Attitude-Measurement: A Reply to 'Opinion Attitude Methodology'." *Psychological Bulletin*, LXIII (1946), 570-589.
6. Crespi, L. P. "'Opinion-Attitude Methodology' and the Polls--a Rejoinder." *Psychological Bulletin*, LXIII (1946), 562-569.
7. Edgerton, H. A., Britt, S. H. and Norman, R. D. "Objective Differences Among Various Types of Respondents to a Mailed Questionnaire." *American Sociological Review*, XII (1947), 435-444.
8. Ellis, A. "The Validity of Personality Questionnaires." *Psychological Bulletin*, XLIII (1946), 385-440.
9. Franzen, R. and Lazarsfeld, P. F. "Mail Questionnaire as a Research Problem." *Journal of Psychology*, XX (1945), 239-320.
10. Hancock, J. W. "An Experimental Study of Four Methods of Measuring Unit Costs of Obtaining Attitude Toward the Retail Store." *Journal of Applied Psychology*, XXIV (1940), 213-230.

11. Hubbard, F. W. "Questionnaires, Interviews, Personality Schedules." *Review of Educational Research*, XII (1942), 534-541.
12. Kelley, T. L. *Scientific Method, its Function in Research and in Education*. New York: Macmillan Co., 1932.
13. Koos, L. V. *The Questionnaire in Education*. New York: Macmillan Co., 1928.
14. Lazarsfeld, P. F. "The Use of Mail Questionnaires to Ascertain the Relative Popularity of Network Stations in Family Listening Surveys." *Journal of Applied Psychology*, XXIV (1940), 802-816.
15. Lindsay, F. E. "Questionnaires and Follow-up Letters." *Pedagogical Seminary*, XXVIII (1921), 373-397.
16. McNemar, Q. "Opinion-Attitude Methodology." *Psychological Bulletin*, XLIII (1946), 289-374.
17. Moore, C. C. "Increasing the Returns from Questionnaires." *Journal of Educational Research*, XXXV (1941), 139-141.
18. Mudrow, L. "Psychologische Studien bei Fragebogenerhebung." *Zeitschrift für Angewandte Psychologie* LI (1936), 337-352.
19. National Education Association, Research Division. "The Questionnaire." *N.E.A. Research Bulletin*, VIII (1930), 1-51.
20. Pace, C. R. "Factors Influencing Questionnaire Returns from Former University Students." *Journal of Applied Psychology*, XXIII (1939), 388-397.
21. Patterson, A. C. "The Questionnaire as a Means of Educational Research. I. The Extent and Reliability of Questionnaire Investigation. II. Sampling the Questionnaire. III. The Construction and Administration of a Questionnaire." *Scottish Educational Journal*, XXV (1942), 683, 698, 708.
22. Ram, P. "On the Questionnaire Method in Psychology." *Indian Journal of Psychology*, XIV (1939), 75-80.
23. Reid, S. "Respondents and Non-Respondents to Mail Questionnaires." *Educational Research Bulletin*, Ohio State University, XXI (1942), 87-96.
24. Reuss, C. F. "Differences Between Persons Responding and Not Responding to a Mailed Questionnaire." *American Sociological Review*, VIII (1943), 433-438.
25. Rollins M. "The Practical Use of Repeated Questionnaire Waves." *Journal of Applied Psychology*, XXIV (1940), 770-772.
26. Shuttleworth F. K. "A Study of Questionnaire Technique." *Journal of Educational Psychology*, XXII (1931), 652-658.
27. Shuttleworth, F. K. "Sampling Errors Involved in Incomplete Returns to Mail Questionnaire." *Psychological Bulletin*, XXXVII (1940), 437.
28. Sletto, R. F. "Pretesting of Questionnaires." *American Sociological Review*, V (1940), 193-200.

29. Stanton, F. "Notes on the Validity of Mail Questionnaire Returns." *Journal of Applied Psychology*, XXIII (1939), 95-104.
30. Suchman, E. A. and McCandless, B. "Who Answers Questionnaires?" *Journal of Applied Psychology*, XXIV (1940), 758-769.
31. Symonds, P. M. *Diagnosing Personality and Conduct*. New York: Century Co., 1931.
32. Thorndike, E. L. *Educational Psychology. Vol. I: The Original Nature of Man*. New York: Teachers College, Columbia University, 1913.
33. Toops, H. A. "Validating the Questionnaire Method." *Journal of Personnel Research*, II (1923), 153-161.
34. Toops, H. A. "The Returns from Follow-up Letters to Questionnaires." *Journal of Applied Psychology*, X (1926), 92-101.
35. Toops, H. A. "Predicting the Returns from Questionnaires: A Study in the Use of Qualitative Data." *Journal of Experimental Education*, III (1935), 204-215.
36. Toops, H. A. "The Factor of Mechanical Arrangement and Typography in Questionnaires." *Journal of Applied Psychology*, XXI (1937), 225-229.
37. Weaver, H. G. "Consumer Questionnaire Technique." *American Marketing Journal*, I (1934), 115-118.

PERSONALITY FACTORS IN THE PSYCHOLOGICAL WEANING OF COLLEGE WOMEN¹

ARTHUR W. SHERMAN, JR.

Ohio University

THIS study is a continuation of an earlier study (2) the purpose of which was (a) to develop an inventory that would be suitable for measuring the emancipation status of older adolescents, and (b) to study certain factors which seem to be related to the emancipation status. In the former study the results obtained by use of the inventory showed that college women were less emancipated than college men.

The purpose of the present study is (a) to determine the relationship between the degree of emancipation of college women and certain personality traits as measured by the *Bernreuter Personality Inventory*, and (b) to determine the relationship between the emancipation status of college women and certain personality traits of their parents.

The *Emancipation Questionnaire* (copyrighted as the *Parental Relationships Inventory*), devised for and used in the original study was revised so that in its present form it includes the 35 items² that were found to be most diagnostic by application of the criterion of internal consistency. An item analysis of the revision confirms the findings from the first study with regard to the diagnostic power of these items. No item was used in the present scale that failed to reveal differences between the most and least emancipated persons. The seven items that have the greatest diagnostic power are:

¹ This study was made possible through a grant from the George Davis Bivin Foundation, Inc.

The writer wishes to express his appreciation to Dr. Irma E. Voigt, Dean of Women of Ohio University, for her cooperation in the securing of the data, and to Dr. James R. Patrick, Professor of Psychology, Ohio University, for certain suggestions in the interpretation of the data.

This paper was read at the annual meeting of the Midwestern Psychological Association in Chicago on May 3, 1947.

² See (2), pp. 174-177, 180. The items included in the revised questionnaire are: 1, 3, 4, 7, 8, 10, 12, 13, 14, 16, 17, 20, 21, 28, 29, 30, 32, 33, 35, 36, 37, 39, 40, 44, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60.

When I am feeling blue I usually look to my parents for comfort.

I often wish that my parents were living nearer to the university so that I could see them more frequently.

After being with my parents for a vacation period, I find it very difficult to leave them.

I often wish that my parents were here at the university so that I might discuss my problems with them.

When writing to my parents, I tell them everything I have done in the past few days.

I often wish that my parents would visit me at the university more frequently.

I decide what courses I shall pursue without assistance from my parents.

Procedure

The revision was administered to 504 college women meeting the following criteria: (a) unmarried (b) not living at home (c) grew to maturity in a home which included the presence of both parents (d) both parents living (e) not a veteran of World War II. The 504 women ranged in ages from 17 years 0 months to 25 years 7 months, the mean age being 19 years 6 months and the median age 19 years 4 months. The women represented the four undergraduate classes at Ohio University as follows: 221 freshmen, 135 sophomores, 92 juniors, and 56 seniors.

The Emancipation scores ranged from 8 (least emancipated) to 33 (most emancipated) out of a possible total score of 35.³ The mean Emancipation score for the group was 21.08 with an S.D. of 5.29.

Since the *Parental Relationships Inventory* was administered in class rooms and to groups in residence halls, sororities, and cottages, there was some duplication. Sixty-two (62) women were retested, but in no case less than two weeks after the first administration. The test-retest reliability coefficient was found to be $.82 \pm .04$.

On the basis of the total scores on the *Parental Relationships Inventory*, the upper 20% (most emancipated) and the lower 20% (least emancipated) were determined. The individuals in these two groups (106 in the most emancipated group and

³ Each emancipated response is given a score of 1. Unemancipated and question-mark responses are given a score of 0.

102 in the least emancipated group) were requested to report at scheduled times for the purpose of accomplishing the *Bernreuter Personality Inventory*. Ninety per cent (90%) of the women scoring in the upper 20% and 90% of the women scoring in the lower 20% completed the Bernreuter.

At the close of the testing period, these individuals were requested to address an envelope to their parents. In this envelope was enclosed a letter prepared by the writer asking the mother and father to accomplish independently and anonymously the Bernreuter Inventory and the Strecker questionnaire.⁴ The forms were clearly marked "Mother" and "Father" and were coded so that they could be paired and separated into the most and least emancipated groups. Replies were received from 82% of the mothers and 75% of the fathers of the most emancipated women who had completed the Bernreuter; 67% of the mothers and 64% of the fathers of the women in the least emancipated group that had completed the Bernreuter replied.

The Findings

The results obtained on the *women* in the least and most emancipated groups are shown in Table 1.

1. It will be recalled that a low score on the B1-N scale is desirable.⁵ This indicates a tendency to be very well balanced emotionally. The difference between the mean raw scores for the most emancipated and least emancipated groups is 31.6, with the most emancipated group having the more favorable score. This difference is significant at the .01 level.⁶

2. On the B2-S scale persons scoring high prefer to be alone, rarely ask for sympathy or encouragement, and tend to ignore

⁴ In private communications between the writer and Dr. Edward A. Strecker, author of *Their Mothers' Sons*, and the publishers, J. B. Lippincott Co., permission was granted for the reproduction of the questionnaire published in Chapter 19 of the above cited book. This questionnaire is essentially a check list for parents to indicate the extent to which they assist the child in obtaining The findings from the administration of this questionnaire further analyses are made. It appears from the manner in which the questionnaire was used in this study, that both the mother and the father in most cases collaborated in answering the items and that they were motivated to reveal themselves in the most favorable light rather than to evaluate their own feelings and actions honestly and objectively.

⁵ These interpretations are taken from Robert G. Bernreuter, *Manual for the Personality Inventory*, Stanford University Press, Stanford University, California, 1935.

⁶ Significance at the .01 level is implied in subsequent references to the significance of differences between mean scores

the advice of others. The difference between the mean scores for the extreme groups is 50, with the most emancipated group being more self-sufficient. This difference is significant.

3. On the B3-I scale a low score indicates the tendency toward extroversion. The difference between the mean scores for the extreme groups is 11.2, with the most emancipated group being more extroverted than the least emancipated group. This difference does not meet the criterion for significance.

TABLE 1
Comparison of Mean Raw Scores on Six Scales of the Bernreuter Personality Inventory Between College Women in the Most and Least Emancipated Groups

Scale	Bernreuter Norms (College Women) Oct. 1938	Most Emanci- pated Group N 95	Least Emanci- pated Group N 93	Diff	s Diff	Diff s Diff	Chances in 100
B1-N							
Mean.....	-42.8	-83.6	-52.0	31.6	11.13	2.84	100 (99.74)
S.D.....	76.8	74.6	80.5				
B2-S							
Mean.....	6.8	25.9	-24.1	50.0	7.85	6.37	100 (99.99)
S.D.....	54.2	56.1	51.3				
B3-I							
Mean.....	-14.7	-43.1	-31.0	11.2	6.68	1.68	96
S.D.....	47.8	44.2	47.0				
B4-D							
Mean.....	30.6	71.8	30.4	41.4	8.71	4.75	100 (99.99)
S.D.....	61.8	57.3	61.0				
F1-C							
Mean.....	8.7	-50.0	1.0	51.0	11.97	4.26	100 (99.99)
S.D.....	75.1	79.3	84.5				
F2-S							
Mean.....	-31.1	-14.1	-68.6	54.5	8.08	6.75	100 (99.99)
S.D.....	56.4	60.2	50.1				

4. On the B4-D scale a high score indicates the tendency to dominate others in face-to-face situations. The difference between the mean raw scores for both groups is 41.4, with the most emancipated group being more dominate than the least emancipated group. This difference is significant.

5. On the F1-C scale a low score indicates the tendency to be wholesomely self-confident and well adjusted. The difference between the mean raw scores for both groups is 51, with the most emancipated group being more self-confident than the least emancipated group. This difference is significant.

6. On the F2-S scale a low score indicates the tendency to be sociable and gregarious. The difference between the mean raw scores for the most emancipated and least emancipated groups is 54.5, with the least emancipated group being decidedly more sociable and gregarious than the most emancipated group. This difference is statistically significant.

It will be noted in Table 1 that the mean raw scores for the most emancipated and least emancipated groups on the B1-N and B3-I scales are below the mean raw scores of Bernreuter's norms, whereas on the B2-S and F2-S scales the most emancipated group lies on one side of the Bernreuter norm group and

TABLE 2

Comparison of Mean Raw Scores on Three Scales of the Bernreuter Personality Inventory Between the Mothers of the College Women in the Most and Least Emancipated Groups

Scale	Bernreuter Norms (Adult Women) Oct. 1938	Mothers of Most Emanci- pated Women N 78	Mothers of Least Emanci- pated Women N 62	Diff	σ Diff	$\frac{\text{Diff}}{\sigma \text{ Diff}}$	Chances in 100
B1-N							
Mean	-27.6	-41.9	-39.2	2.7	14.91	.18	57
S.D.	79.2	86.2	88.7				
B2-S							
Mean	3.2	10.1	6.8	3.3	10.46	.31	62
S.D.	56.8	62.7	60.5				
B4-D							
Mean	18.4	29.5	24.0	5.5	10.78	.51	69
S.D.	63.0	63.0	63.6				

the least emancipated group on the other side, as might be expected to occur. For the B4-D and F1-C scales we find that the mean scores for the least emancipated group lie very close to the mean scores for the Bernreuter norm group.

The results from the Bernreuter Inventory on the *parents* of the women in the most emancipated and least emancipated groups are shown in Tables 2 and 3. The inventories received from the parents were carefully checked to determine the extent to which parents might have collaborated in completing the inventory. It was found in checking the ten cases where the raw scores for the mother and father were within ten points of each other that this proximity was not the result of an item-by-item agreement on the Bernreuter.

In Table 2 is shown the comparison of the mean raw scores on the B1-N, B2-S, and B4-D scales of the Bernreuter between the *mothers* of the women in the most and least emancipated groups. The differences found are very slight and statistically insignificant.

In Table 3 is shown the comparison of mean raw scores on the B1-N, B2-S, and B4-D scales of the Bernreuter between

TABLE 3
Comparison of Mean Raw Scores on Three Scales of the Bernreuter Personality Inventory Between the Fathers of the College Women in the Most and Least Emancipated Groups

Scale	Bernreuter Norms (Adult Men) Oct. 1938	Fathers of Most Emanci- pated Women N 73	Fathers of Least Emanci- pated Women N 79	Diff	s Diff	Diff s Diff	Changes in 100
B1-N							
Mean	-63.9	-68.6	-72.0	4.1	15.88	25	60
S.D.	79.0	100.6	81.9				
B2-S							
Mean	32.3	41.6	27.9	13.7	8.3	165	95
S.D.	50.0	45.5	48.9				
B4-D							
Mean	53.9	66.0	64.1	1.9	10.30	18	57
S.D.	58.0	67.3	51.7				

the *fathers* of the women in the extreme groups. The differences here are likewise very slight and/or statistically insignificant.

Other Factors Related to Emancipation Status

The most emancipated group was found to be somewhat more intelligent, on the average, than the least emancipated group. A significant difference of 10.98 was found between the two groups on the basis of raw scores on the *Ohio State Psychological Examination* (Forms 21 and 22). This corroborates the findings from the original study (2, p. 179).

The most emancipated group was found to be 8.9 months older, on the average, than the least emancipated group. This corroborates the findings from the original study (2, p. 179).

In this study the ages and schooling of the mothers and fathers were studied for one-half of the group. These factors seem not to be related to emancipation.

Summary

In summary, we find these statistically significant differences between the most emancipated and the least emancipated groups:

1. The most emancipated women were found to be more stable emotionally, on the average, than the least emancipated women.

2. The most emancipated women were found to be considerably more self-sufficient, on the average, than the least emancipated women.

3. The most emancipated women were found to be more dominate, on the average, than the least emancipated women.

4. The least emancipated women were found to be considerably more sociable and gregarious, on the average, than the most emancipated women. This finding suggests the possibility that friends and associates may in some respects play a role similar to that of the parents, and the feeling of helplessness may be the motivating mechanism that "moves her toward people."⁷ If this discovery proves to be true then one needs to examine more carefully the motives of *why* people seek one another. Sociability alone may not be a sign of maturity.

5. The most emancipated group is, on the average, 8.9 months older than the least emancipated group. This corroborates the findings from the original study.

6. The most emancipated group is, on the average, somewhat superior in intelligence to the least emancipated group. This likewise corroborates the findings from the original study.

7. There is no appreciable or significant difference between the two groups on the B3-I (introversion-extroversion) scale.

8. The differences found between the mothers of the most emancipated women and the mothers of the least emancipated women on the B1-N, B2-S, and B4-D scales of the Bernreuter are very slight and statistically insignificant.

9. The differences found between the fathers of the most emancipated and least emancipated women are very slight,

⁷ Horney (1) categorizes the development of conflicts in children into "moving toward people, against them, or away from them" (p. 42). "When moving toward people he accepts his own helplessness, and in spite of his estrangement and fears tries to win the affection of others and to lean on them."

except for the B2-S scale, with a difference of 13.7. None of these differences, however, are significant.

10. The ages and schooling of the parents were found, on the average, to bear no relation to the emancipation status of the college women studied.

All of the above statements hold insofar as we can accept Bernreuter's interpretations of his scales and if our scale proves to be a valid measure of emancipation.

This problem requires further study of the relationship of the adolescent to the parents and the methods of control employed by the parents.* Detailed analysis of the interacting personalities of parents and child is essential to an understanding of the process of psychological weaning and the achievement of emancipation by the older adolescent.

REFERENCES

1. Horney, Karen. *Our Inner Conflicts*. New York: W. W. Norton and Company, 1945.
2. Sherman, Arthur W., Jr. "Emancipation Status of College Students." *Journal of Genetic Psychology*, LXVIII (1946), 171-180.
3. Strecker, Edward A. *Their Mothers' Sons*. Philadelphia: J. B. Lippincott Co., 1946.

* The writer is investigating this problem further.

NOTE CONCERNING "THE VALIDITY OF STANDARD
AND CUSTOM-BUILT PERSONALITY INVENTORIES
IN A PILOT SELECTION PROGRAM," BY DONALD
E. SUPER.

DONCASTER G. HUMM

Los Angeles, California

DONALD E. Super's study of the applicability of personality inventories reports such a lack of effectiveness on the part of several standardized personality tests that his article needs critical consideration.

Since personality inventories have an important potential contribution to make to our understanding of human behavior and to our ability to predict the adjustment of individuals to various situations, they deserve extensive examination made carefully and with due regard for their complexities.

Dr. Super states that three standard personality inventories, the *Shipley Personal Inventory*, the *Adams-Lepley Personal Audit*, and the *Humm-Wadsworth Temperament Scale*, have "no significant relationship with success in flying training." He mentions the Guilford Inventories, the *Maller-Glaser Interest-Values Inventory*, and the *Minnesota Multiphasic Personality Inventory* in passing, but reports that they were not used.

So far as the *Humm-Wadsworth Temperament Scale* is concerned, the study so briefly reported by Dr. Super does not justify any conclusion whatever as to its validity or as to any contribution it might have made to the selection of candidates for pilot training. An understanding of the circumstances may be helpful.

At the request of an officer of the Air Corps who had been authorized in his civilian position to use the *Humm-Wadsworth Temperament Scale* (not Dr. Super), I supplied Humm-

¹ This article by Donald E. Super appeared in *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VII (1947), 735-744.

Wadsworth Temperament Scales and scoring and interpreting materials for administering the Scale to 500 cadets in pilot training. I did so with the understanding that I was to participate in the study, although the nature and extent of my participation were not formally defined. Thereafter, I was unable to secure any response to inquiries as to how the Scale was being administered and interpreted or what statistical procedures were to be used in studying the results of its use or what criteria were to be used in evaluating it, except the statement that all this was a military secret.

Inasmuch as I knew that the persons conducting the experiment did not have the specialized knowledge needed to handle the Temperament Scale, I appealed to the Adjutant-General's office. An order was issued that the use of the *Humm-Wadsworth Temperament Scale* should be discontinued. However, another ranking officer intervened, and I was requested to permit the experiment to continue and to give assistance to it. I was happy to do this. At that time, I was assured that the results of the experiment would be made available to me and that I was to have an opportunity to know what procedures were to be used. I received no further information.

Instead, I received a listing of the scores of 382 aviation cadets, identified by number only, from which we prepared a description of the temperament of each subject. These were returned to the unit with the warning that the data were insufficient for prediction of success in flying, except in some exceptional instances. Their receipt was not acknowledged, and no report of the outcome of training of the subjects was given to me.

Of the 382 subjects whose scores were submitted to me, 201 showed measures of response-bias within the probably acceptable range. Since Dr. Super states that the study he reports was based on 202 cases, I infer that this study was limited to the cases whose response-bias was within the range of probable acceptability.

It is evident from Dr. Super's report that the interpretations of temperament I made were not used and the warnings not observed. Instead, correlations were computed between the

scores of the separate components and a criterion—graduation-elimination in primary pilot training.

This is an erroneous procedure because the scores in seven components of the *Humm-Wadsworth Temperament Scale* are meaningful only when considered in combination

The most simple combination which can be used at all in predicting behavior is the relationship between the "Normal" score and the score in each of the other components. Even this represents a serious oversimplification, since the other components also tend to reinforce each other or to inhibit each other or to give rise to unresolved conflict (a frequent occurrence which explains many cases of maladjustment and apparently inconsistent behavior). For example, the chief difference between the average temperament of successful workers and the average temperament of patients with manic-depressive psychoses is in the amount of integration present as indicated by the relationship between the Normal and the Cycloid components. Successful workers are usually acceptably integrated; manic-depressive patients are, for the duration of their illness, disintegrated. Thus there is a vast difference between strong Manic accompanied by strong Normal and strong Manic accompanied by weak or borderline Normal. Again, strong Manic with strong Autistic presents a very different picture from strong Manic with weak Autistic or strong Manic with strong Paranoid. Again, strong Normal may be overwhelmed by an exceptionally strong other component.

Failure to consider these relationships among the components predisposes Dr. Super's findings to the influence of counter-acting tendencies within the data which would mask the truth.

A second problem which arises is that of response-bias. It is not apparent from Dr. Super's report whether raw scores or profile points were used in the compilation of the data. If raw scores were used, response-bias was not accounted for except by eliminating those cases likely to be most seriously affected by this factor. This, however, would have the effect of curtailing the range and thus of artificially lowering the correlations. If profile points were used, then response-bias was compensated so far as it affected the cases that were in-

cluded, but unduly low correlations because of the effect of curtailed range would still result. Dr. Super could have compensated for curtailment of range by the use of the following formula:

$$r = r' \frac{\sigma_{12}}{\sigma_{(1,2)'}}$$

(In which r = the obtained correlation. r' = the correlation obtained from data which include the full range. σ_{12} is the standard deviation of the estimate as obtained, and $\sigma_{(1,2)'}$ is the standard deviation of the estimate of the full array.)

A third question which arises is that of any attempt to use the correlation coefficient with such data. Two points need to be mentioned. The first, our own acquaintance with these data indicates that the arrays are markedly skewed. Correlation ratio or correlation index would have been preferable to product-moment correlation. The second point is that, in such a problem as this, which represents test validation on remote criteria (1), some form of multiple or partial correlation measures is necessary in order to estimate the influence of other factors present. To accept such a standard as low positive correlation between one of the members of a test battery and success on the job is to leave many factors unaccounted for.

A fourth question arises from the criteria used. Ordinarily, teachers' judgments and supervisors' ratings seldom exceed a validity in which $r = +.50$. Dr. Super makes no mention of the validity of the ratings used. His findings should have been corrected with this formula:

$$\text{From } r_{1a} = \frac{r_{01}}{\sqrt{r_{0a}}}$$

(Validity of a score corrected for the validity of the criterion.)

That the *Humm-Wadsworth Temperament Scale* can be used to measure temperament validly and is useful in situations where prediction of behavior is needed has been demonstrated by a number of studies. One of our own (2) showed agreement between Scale scores and carefully conducted and verified case

studies in 91 per cent of the comparisons (roughly equivalent by the method of unlike signs to a correlation of + 96).

In another instance, 506 war emergency policemen were given an intelligence test and the *Humm-Wadsworth Temperament Scale* during their training period and for each one a prediction of success on the job was recorded on the basis of these two tests. At the end of a year, 103 of these men had been discharged for proved violations of police regulations or of the criminal law. Of these 103, one had been rated "likely to succeed," 16 had been rated "chance of success doubtful," 61 had been rated "likely to fail," and 25 had been rated "very likely to fail." Dividing the 16 predictions of doubtful chance of success in half, we have here accurate prediction in 91 per cent of the cases.

In a third instance, 65 long-term employees of a company were given the Temperament Scale. The forms, identified by number only, were delivered to a technician who, from the Scale results alone, wrote a description of the temperament of each subject. These descriptions were then read to a group of long-term supervisors who attempted to name each subject. They were successful in 62 of the cases (95 per cent, roughly equivalent to a correlation of +.99).

Recently we have reviewed reports from fourteen users of the Scale whose experience covered a period of years. 20,000 cases were included. The weighted average indicates that the Scale afforded accurate reports of temperament in 93.5 per cent of the cases.

We have not received any report of failure of the Scale in any situation where it was used as it was designed to be used and was used with due regard to correct procedures.

REFERENCES

1. Humm, Doncaster G. "Test Validation on Remote Criteria." *Journal of Applied Psychology*, XXX (1946), 333-339.
2. Humm, Doncaster G. and Humm, Kathryn A. "Validity of the Humm-Wadsworth Temperament Scale." *Journal of Psychology*, XVIII (1944), 55-64.

MEASUREMENT ABSTRACTS¹

Ammons, R. B. and Hieronymus, A. N. "Critical Evaluation of a College Program for Reading Improvement." *Journal of Educational Psychology*, XXXVIII (1947), 449-470.

All members of the entering class of 1941 at the University of Iowa were given a battery of five tests designed to measure reading rate, comprehension, etc. One hundred and sixty-seven students scoring below the fiftieth percentile on the *Blommers Reading Test* were given a special twenty-hour reading instruction course. A control group of 42 students were selected from the same population. Upon retesting, the experimental group showed significant gains over the control group in three of the five tests, including the *Blommers Reading Test*. There was a gain in reading rate with no loss of comprehension. However, prediction of gain for individual students could not be made from the results of the study. The authors conclude that twenty hours of training can result in substantial gains in 'free' reading. *Dorothy Stock*

Billingslea, Fred Y. "The Bender-Gestalt: An Objective Scoring Method and Validating Data." *Journal of Clinical Psychology*, IV (1948), 1-27.

Assuming the original premise of the Bender-Gestalt test, that accurate visual-motor perceptual behavior is a skillful act involving sensory reception, central neural interpretation, and motor reproduction, which may be distorted by neural injury, by variations in the intellectual level, and by maladjustments in the emotional organization, this study sees as its problem the quantitative evaluation of the amount of distortion in the figure reproduction, the discovery of the patterns which the distortions take, and the relating of these to the disturbances judged to be present in the subject's personality organization. The experimental group consisted of 100 psychoneurotic adult males; the control group of 50 males equated in intelligence, occupational experience, etc. Sixty-three indices were established to give quantifying coefficients to 25 test factors. No clear-cut reliability for the instrument was found. As regards validity, it is concluded that a clear syndrome for distinguishing the psychoneurotic test record has not been established, but that the

¹ This is the last Measurement Abstracts section to be edited by Professor Forrest A. Kingsbury, who retires from his professional duties at the University of Chicago at the end of the present semester. Professor Kingsbury has given enthusiastic and faithful support to *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT* ever since its conception in 1940, and he has made a substantial contribution to its usefulness. He leaves the staff with the sincere thanks and hearty best wishes of the rest of us.

G. F. K.

tool can be of use for intuitive judgments of the client's clinical behavior. Suggestions for improved standardization and administration are indicated. *Granville C. Fisher.*

Bradley, N. C. "The Growth of the Knowledge of Time in Children of School Age." *British Journal of Psychology*, XXXVIII (1947), 67-78.

The aim of this study was to trace in children the growth of the ability to understand ordinary time-words and the development of a concept of the universal and continuous nature of the time-scheme. Results are reported for 4 tests administered to urban school children aged 5 to 13, and indicate a general order of development. Distinctions between the present, past, and future are found to be established by the age of 5; the first time-words to be used following this stage refer to natural phenomena and personal activities or relationships, after which understanding is achieved of words more purely conventional in nature. Capacity to understand the conventional time-scheme appears to be developed later than is generally believed, and this observation is of educational significance, particularly in the teaching of history. *Frances Smith.*

Brown, Andrew W., Morrison, Joan and Couch, Gertrude B. "Influence of Affectional Family Relationships on Character Development." *Journal of Abnormal and Social Psychology*, XLII (1947), 422-428.

The subjects for this study are a group of ten-year-olds and a group of sixteen-year-olds in a small midwestern town. The first group was retested after three years according to a revised system. Questionnaires were developed to measure affectional family relationships and character reputation. The correlations between these two factors show that affectional family relationships determine to a fair degree the type of personality a child will possess. The area of family relationships most closely related to character development is Sharing in Family Decisions (the degree to which a democratic atmosphere prevails in the home). Also important are Interparental Relations and Parental Attitude toward Peer Activities. *Dorothy Stock.*

Bruner, Jerome S. and Postman, Leo. "Emotional Selectivity in Perception and Reaction." *Journal of Personality*, XVI (1947), 69-77.

Nineteen subjects were presented with 99 stimulus words, some of which were potentially emotional, and their reaction time was recorded. A second list was made up for each subject which included the six words yielding the fastest associative reaction time, the six yielding the slowest reaction time, and the middle six words. This list, presented two weeks after the first list, was used to measure speed of reaction. The results show a curvilinear relationship

between the two distributions. The authors believe that two functions combine to produce the results: an intensification of defense in the face of increasing emotionality (where a slow associative reaction time is accompanied by a slow rate of recognition), and a sensitization to tension producing stimulus objects (where a slow associative reaction time is accompanied by a very fast rate of recognition). The results show the equivalence of perceptual defense and associative blocking. *Dorothy Stock.*

Cattell, Raymond B., and Luborsky, Lester B. "Personality Factors in Response to Humor." *Journal of Abnormal and Social Psychology*, XLII (1947), 402-421.

Believing that some of the more profound aspects of a man's personality may be revealed by observing the things at which he laughs most heartily, the authors begin in this study an attempt to ascertain the constancy of humor appreciation, the primary tendencies to be found expressed in humor reactions, the effect of familiarity with the jokes, and (reserved for a future study) the relation of the measured humor responses to measured personality factors and personal background. Using the responses of 50 men and 50 women to 100 selected jokes, 13 well-defined clusters appeared, an analysis of which resulted in five factors, believed to be general personality factors. An attempt at serious interpretation of these factors is considered premature, but it is felt that they may prove to be of clinical and research value in the light of further study. *Granville C. Fisher.*

Cockett, R. "The Nature of Performance Tests: An Experimental Study." *British Journal of Psychology*, XXXVIII (1947), 29-38.

The author performed four experiments (testing subjects' performance on specified manual tasks) in an attempt to find what type of activity a performance test requires in order to be reliable in the psychometric sense. His first hypothesis was that a performance test requires some minimum degree of complexity before it can be reliable. Experiments 1 and 2 were performed to test this hypothesis, but the hypothesis was not substantiated by the results. Experiments 3 and 4 were then performed and the results yielded the conclusion that "a fundamental requirement in a performance test is that it must demand genuinely *psychophysical* activity throughout, i.e. . . . continuous conscious guidance of bodily action." *Sarah Counts.*

Deemer, Walter L., Jr. "The Power of the *t* Test and the Estimation of Required Sample Size." *Journal of Educational Psychology*, XXXVIII (1947), 329-342.

The purpose of the paper was to review the concept of the 'power' of a statistical test and to develop a formula for use in determining the required sample size to achieve a desired power of the *t* test. Use of the *t* test may result in either one of two kinds of errors: (1) re-

jection of the null hypothesis when it is true; (2) acceptance of the null hypothesis when it is false. An 'error of the first kind' is guarded against by employing the usual significance levels. A method is presented for estimating, after an experiment has been performed, the probability of having committed an 'error of the second kind.' Another method is given for estimating in advance the sample size required to attain any given probability of rejecting the null hypothesis when the probabilities of committing both kinds of errors and the difference between means are fixed. Both symmetrical and asymmetrical *t* tests are considered. *Kay C. Montgomery.*

Ellis, Albert. "A Comparison of the Use of Direct and Indirect Phrasing in Personality Questionnaires." *Psychological Monographs*, No. 284, 35 pp.

Thirty-six most frequently used questions on eight standard personality questionnaires for children comprise the basic personality test used in this study. Two direct forms and two indirect forms of the 36 items were constructed, and administered to 47 behavior-problem boys and 221 non-problem boys. There were three orders of presentation: indirect-direct, alternate, and direct-indirect. The results suggest that (1) in order to encourage truthful responses indirect items should precede direct items; (2) each new mode of indirectly phrasing a series of questions may result in a significantly different instrument; (3) indirect tests do not differentiate problem from non-problem groups any better than direct tests do, but individual items may be more discriminating when indirectly phrased than when directly phrased. *Dorothy Stock.*

Fassett, Katherine K. "A Preliminary Investigation of the Sargent Test." *Journal of Clinical Psychology*, IV (1948), 45-56.

Twenty-five women students from beginning psychology classes were given the original (S-form) of the *Sargent Test* and 25 were given the experimental (X-form) of the test. A re-test was administered to both groups six weeks later. The S-forms were scored according to the original Sargent method; the X-forms according to the simplified experimental method. The author found that the test-retest reliability of the *Sargent Test* is reasonably high for a projective technique. The X-form gives a larger sampling of conflict areas while the S-form gives a larger number of questions in the replies. The author concludes that the experimental scoring system has potential value as a clinical device. *Dorothy Stock.*

Fiske, Donald W. "Validation of Naval Aviation Cadet Selection Tests Against Training Criteria." *Journal of Applied Psychology*, XXXI (1947), 601-614.

The three tests, the *Personnel Test* devised by Wonderlic from the *Otis S-A Test*, the *Mechanical Comprehension Test* constructed for the Navy by Bennett, and the *Biographical Inventory* developed by Kelly,

used to screen applicants for Navy flight training, are reported on in terms of the correlations between each test and outcome of training. The *Personnel Test* was found useful in predicting ground school failures. The *Mechanical Comprehension Test* identified flight and ground school failures. The *Biographical Inventory* was best at predicting flight failures. *Granville C. Fisher.*

Fox, Charlotte. "Vocabulary Ability in Later Maturity." *Journal of Educational Psychology*, XXXVIII (1947), 482-492.

Word lists chosen from the *English Recognition Vocabulary Test* of Seashore and Eckerson were administered to 30 subjects aged 70 to 79, and to a control group of 30 subjects aged 40 to 49. Three methods of testing were employed: group recognition (multiple-choice), individually administered oral recognition, and oral definition. No significant differences were found between the two age groups, the definition task being significantly more difficult than the individual recognition task for both groups. Statistical analysis of half-credit definition scores for the two groups showed no significant difference in quality of definitions given by the older and younger subjects. *Frances Smith.*

Hsü, E. H. "The Rorschach Responses and Factor Analysis." *Journal of General Psychology*, XXXVII (1947), 129-138.

Detailed records were made of the Rorschach responses of 76 children with normal IQ's and an age range of 5 to 15. Plate I was selected as offering the largest number of common responses, and 15 variables were chosen on the basis of responses to this card, for computation of intercorrelations. The variables included number of words, number of nouns, of adjectives, and of verbs, use of number, and total time, as well as specific common content-categories. Factor analysis by the centroid method and rotation of axes gave a factor matrix of 7 columns. The psychological meaningfulness of 6 of the factors is discussed. *Frances Smith.*

Jolles, Isaac. "The Diagnostic Implications of Rorschach's Test in Case Studies of Mental Defectives." *Genetic Psychology Monographs*, XXXVI (1947), 89-198.

Sixty-six children whose psychometric records indicated seriously retarded mental development were subjected to further study by means of their case histories and use of the *Rorschach Test*. On the basis of subsequent findings, they were divided into 3 groups: those whose Rorschach patterns clearly indicated normal mental ability, those whose patterns were inconclusive, and those definitely retarded. In comparison of the 3 groups an extensive overlapping in IQ's was found. In each of the 66 cases Rorschach results showed severe emotional disturbance, while case history data revealed the importance of family background in development of mental deficiency. It is emphasized that psychotherapeutic techniques as well as special

educational facilities should be utilized in treatment of the mentally deficient, and that psychometric examination alone does not provide data for adequate study of the retarded child. *Frances Smith.*

Knauff, Edwin B. "A Classification and Evaluation of Personnel Rating Methods." *Journal of Applied Psychology*, XXXI (1947), 617-625.

In view of a felt need for more precise methods of evaluating the performance of workers, together with a lack of available psychometric precision in the rating methods currently in use, the author presents a classification of rating methods on the basis of (1) the operations performed by the experimenter or scale maker in constructing the scale; (2) the operations performed by the judges or raters when applying the rating device to a given individual; and (3) the operations involved in constructing a scoring method for the rating device, for the purpose of evaluating the various rating methods and understanding the assumptions underlying their construction and use. The aim is that of stimulating a more critical approach toward the many rating systems now in use. *Granville C. Fisher.*

Leeds, Carroll H., and Cook, Walter W. "The Construction and Differential Value of a Scale for Determining Teacher-Pupil Attitudes." *Journal of Experimental Education*, XVI (1947), 149-159.

The authors report the construction of an attitude scale, the *Teacher-Pupil Inventory*, for gauging teacher attitudes toward pupils and differentiating those teachers who get along well with pupils from those who do not. The inventory consists of 164 items, and requires 20-30 minutes to administer. Its validity, as determined by correlating inventory scores with the combined ratings of the principal, the author, and the pupils, was .59 for a group of 100 unselected teachers; its reliability was .91. Data are presented on the relationship between "... the attitudes of teachers toward pupils and certain personal data relating to the teacher ..." and on "... an analysis of pupil's reactions toward particular teachers." *Kay C. Montgomery.*

Lindgren, Henry C. "A Study of Certain Aspects of the Lee-Thorpe Occupational Interest Inventory." *Journal of Educational Psychology* XXXVIII (1947), 353-362.

Some aspects of the use and possible interpretation of the Lee-Thorpe *Occupational Interest Inventory* were studied. The possible limitations of the Inventory are discussed by the author and he suggests that perhaps its best use is as an interviewing aid. The scores obtained by adult males on the Inventory (advanced Series) were correlated with their scores on the *Kuder Preference Record* and the *Gamma Test*, *Otis Quick-Scoring Tests of Mental Ability*. Significant correlations were found between the Kuder and the Inventory on comparable categories. The correlations between the Otis Gamma and the various scales of the Inventory were low, with the exception

of one—Level of Interests on the Inventory—which approached significance. *Sarah Counts.*

Loevinger, Jane. "A Systematic Approach to the Construction and Evaluation of Tests of Ability." *Psychological Monographs*, LXI (1947), No. 4.

This study is concerned with the problem of the consistency of a test, the choice of subtests or items, and the unit of measurement, as these are related specifically to power tests of mental ability. Reviewing the concept of reliability, the author concludes that as presently defined and used it is highly unsatisfactory. The problem of item selection is discussed and current practices are criticised. Thorndike's, Thurstone's and Flanagan's methods of scaling are evaluated. Developing the thesis that the solution to the above is inherent in the concept of homogeneity, the author presents a series of theorems and proofs, together with formulae for the measurement of homogeneity in a test, and homogeneity in test items. Criteria are offered for an adequate system of scaling. The author states that the subject of scaling is far from closed, no system having proved adequate by the criteria she proposes; the development of adequately scaled tests awaits the development of highly homogeneous tests. *Granville G. Fisher.*

Moore, Thomas V., Stafford, John W. and Hsü, En Hsi. "Obverse Analysis of Personality." *Journal of Personality*, XVI (1947) 10-48.

Fifty-six women students rated themselves on 128 personality traits. Each person was correlated with each of the others in terms of traits commonly claimed or denied, or differentially claimed or denied. This analysis brought out nine factors, three of which could be identified as a cycloid factor, a schizoid factor, and a suspicious factor. A second analysis was undertaken after the 38 least differentiating items were eliminated. This analysis brought out eleven factors: an adjustment factor, a romantic factor, a factor of inferiority, a factor that is not clear-cut, an extrovert factor, a frustration-romantic factor, a schizoid factor, a pre-psychotic "paranoid" factor, a jealousy factor, a factor not clear-cut, possibly an adjusted introvert type, and a pre-psychotic "simple or catatonic schizophrenia" factor. The authors conclude that the first type of analysis is most useful in ascertaining the homogeneity of a population, the second in analyzing configurations of traits. *Dorothy Stock.*

Mummery, Dorothy V. "An Analytical Study of Ascendant Behavior of Pre-school Children." *Child Development*, XVIII (1947), 40-81.

Seventy-nine categories of ascendant behavior were classified according to social acceptability. Forty-two pairs of three- and four-year-old children were observed and rated. Seven of the twelve least

ascendant children were given training in mastering toys; the remaining five constituted a control group. A re-test showed that there was an increase in the total ascendant behavior and in the socially acceptable ascendant behavior for the experimental group while the control group showed no consistent change. The author concludes that encouraging an attitude of confidence in the child will lead to an increase in socially acceptable ascendant behavior without undue reference to the status of the other child. *Dorothy Stack.*

Porteus, Stanley D., and Peters, Henry N. "Maze Test Validation and Psychosurgery." *Genetic Psychology Monographs*, XXXVI (1947), 3-86

Fifty-five patients who had undergone lobotomy at the Kaneshe Hospital for the Insane, Island, were divided into 3 groups on the basis of unimproved, improved, and greatly improved post-operative adjustment. Results for each group of pre-operative and post-operative testing with the *Porteus Maze Test* were compared. A significant decline in performance was found for all patients. When post-operative testings were repeated, the 2 improved groups, though showing the greatest initial decline, were found to make larger subsequent gains, with the much improved group finally excelling the middle group. Whereas Binet-measured functions have shown little change following lobotomy, the Maze Test conclusively shows sensitivity to changes in the functions of planning and foresight. *Frances Smith.*

Shaffer, Robert H., "The Effect of English Deficiency Upon a Student's Adjustment in College." *Bulletin of the School of Education, Indiana University*, XXIV, 1 (1948), 34 pp.

In this study students with a deficiency in English are compared with a matched group without such deficiency in respect to scholastic, social and emotional adjustment. It was found that the deficient students passed significantly fewer hours of work and earned significantly fewer credit points than the non-deficient, the differences between the two groups decreasing steadily from the first to the fifth semester. On the basis of the *Bernreuter Personality Inventory*, the first semester deficient students were found to be (a) less neurotic and more evenly balanced emotionally than the first semester non-deficients, (b) more extroverted and less apt to worry, (c) more dominant, (d) more self-confident and better adjusted to their environment, and (e) more sociable and gregarious. The non-deficient group became more extroverted, more self-confident, and more sociable than college students as a whole, while the deficient group became less dominant, less confident, and less sociable as they remained in college. More of the deficient than the non-deficient first-semester students were found to be participating in group activities. No difference was found between the two groups with regard to (1) dates, dances, and shows; (2) vocational choice having been made; (3) student's own evaluation

of his adjustment; (4) personal or family problems, (5) idea and conception of a successful college; etc. The author offers some interpretations and implications of the data. *Granville C. Fisher.*

Strong, Edward K., Jr. "Norms for Graduate School Business Students on the Minnesota Vocational Test for Clerical Workers." *Journal of Applied Psychology*, XXXI (1947), 594-600.

In view of the erroneous interpretations which are made of the scores of the above test due to a lack of norms applicable to business men or students headed for business rather than clerical positions, the author supplies norms on 168 male business students. He also supplies data to indicate that the test measures ability to see differences in numbers and names, rather than "the ability to discriminate small differences rapidly," as is claimed for the test. Additional data suggest that there is a significant positive correlation between the test and intelligence except among a homogeneous group of clerical workers for whom the test is designed. *Granville C. Fisher.*

Tilton, J. W. "The Relation Between IQ and Trait Difference as Measured by Group Intelligence Tests." *Journal of Educational Psychology*, XXXVIII (1947), 343-352.

Data were collected and analyzed for unevenness or scatter on group intelligence tests. The breakdown of the data was done by holding mental age constant and grouping against IQ. Mental ages from 8 to 15 were used with special emphasis on the 10:00-11:11 range. The profiles for the bright children had the least amount of scatter while those of the dull children showed the greatest unevenness. No attempt was made to account for the relationship between IQ and unevenness. *Sarah Counts.*

Van der Merwe, A. B. and Theron, P. A. "A New Method of Measuring Emotional Stability." *Journal of General Psychology*, XXXVII (1947), 109-124.

This study investigates the value of the *Goetz Finger Plethysmograph* in measuring emotional stability. An independent estimate of emotional stability for each of 30 subjects was obtained through the administration of the *Bell Adjustment Inventory*. The correlation between results for the *Bell Adjustment Inventory* and the *Goetz Finger Plethysmograph* was significant above the 5% level. The author concludes that the rate of reflex finger volume change is a valid measure of emotional stability (as tested by the *Bell Inventory*); that the rate of reflex finger volume differentiates between individuals who are emotionally stable and unstable; and that the *Goetz Finger Plethysmograph* is a useful means of measuring emotional stability. *Dorothy Stock.*

Vernon, M.D. "Different Types of Perceptual Ability." *British Journal of Psychology*, XXXVIII (1947), 79-89.

The discovery of apparent discrepancies between sensory acuity and perceptual skill and of variations in perceptual skill with type of task formed the basis for this study. 30 R.A.F. air crew in training were presented by means of both tachistoscopic and non-tachistoscopic techniques with varied perceptual material, in some of the tests a differential brightness threshold also being obtained. The results offered some indication of 2 general factors influencing perception, analogous to those isolated by Thurstone. These are described as (1) the ability to perceive and to discriminate pure shape and pattern characteristics, which appears to be related to the differential brightness threshold, and (2) the ability to make further interpretation of sensory and perceptual data in accordance with a predetermined set or schema. Here, administration of intelligence tests gave results indicating a relationship with such ability. *Francis Smith.*

Weschberg, Florence Orm, and Sparer, Phineas J. "A Statistical and Profile Comparison of a Hospitalized Tuberculous Group with a Normal College Group Using the Cornell Index." *Journal of Clinical Psychology*, IV (1948), 63-69.

This study attempted to ascertain the efficiency of the *Cornell Index* "stop," "neurotic" and total items in terms of their ability to differentiate the personality variances in a hospitalized tuberculous group of 100, from those in a non-hospitalized sample of 100 college freshmen. Using a newly devised technique of profile analysis, the authors found that on the whole, the *Cornell Index*, N.3, does not seem to show significant personality differences between the compared groups; that "neurotic" items appear to be more discriminating than the "stop" questions; that age is not significantly related to the scores; that the length of hospitalization has a low positive relationship with the neurotic scores of the male patient; that female patients score higher than male in most symptom categories. A salient contribution is claimed for the profile technique. *Granville C. Fisher.*

ADDITIONAL ARTICLES NOT ABSTRACTED

Allen, Robert M. "A Note on the Use of the Wechsler-Bellevue Scale Mental Deterioration Index with Brain Injured Patients." *Journal of Clinical Psychology*, IV (1948), 88-89.

Baggaley, Andrew R. "The Relation Between Scores Obtained by Harvard Freshmen on the Kuder Preference Record and Their Fields of Concentration." *Journal of Educational Psychology*, XXXVIII (1947), 421-427.

Baldwin, Alfred L. "Changes in Patent Behavior During Pregnancy." *Child Development*, XVIII (1947), 29-39.

Chester, David J. "The Wonderlic Personnel Test as a Predictor of Scores on the American Council on Education Examination." *Journal of Clinical Psychology*, IV (1948), 82-85.

Davis, Frederick B. "The Interpretation of Principal-Axis Factors." *Journal of Educational Psychology*, XXXVIII (1947), 471-481.

- Gough, Harrison G. "The Frame of Reference of the Thematic Apperception Test." *Journal of Clinical Psychology*, IV (1948), 90-92.
- Gray, Horace. "Jung's Psychological Types: Meaning and Consistency of the Questionnaire." *Journal of General Psychology*, XXXVII (1947), 177-186.
- Greene, Ronald R. "Studies in Pilot Selection: II. The Ability to Perceive and React Differentially to Configurational Changes as Related to the Piloting of Light Aircraft." *Psychological Monographs*, No. 286 (1947), 18-27.
- Hovey, H. Birnet. "Detection of Circumvention in the Minnesota Multiphasic Personality Inventory." *Journal of Clinical Psychology*, IV (1948), 97.
- Koshuk, Ruth P. "Developmental Records of 500 Nursery School Children." *Journal of Experimental Education*, XVI (1947), 134-148.
- Lane, G. Gorham. "Studies in Pilot Selection: I. The Prediction of Success in Learning to Fly Light Aircraft." *Psychological Monographs*, No. 286 (1947), 1-15.
- LeShan, Eda, and LeShan, Lawrence. "Projection Technique in Social Case Work Procedure." *American Journal of Orthopsychiatry*, XVIII (1948), 73-91.
- Macauley, William J. "The Difficulty of Grammar." *British Journal of Educational Psychology*, XVII (1947), 153-162.
- McCurdy, Harold G. "Basal Metabolism and Academic Performance in a Sample of College Women." *Journal of Educational Psychology*, XXXVIII (1947), 363-372.
- McCurdy, Harold G. "Group and Individual Variability on the Goodenough Draw-a-Man Test." *Journal of Educational Psychology*, XXXVIII (1947), 428-436.
- McGann, Mary. "Diagnostic Testing and Remedial Teaching for Common Errors in Mechanics of English Made by College Freshmen." *Journal of Educational Psychology*, XXXVIII (1947), 499-503.
- Moore, Jean K. "Speech Content of Selected Groups of Orphanage and Non-Orphanage Preschool Children." *Journal of Experimental Education*, XVI (1947), 122-133.
- Mullen, Frances A. "The Minnesota Multiphasic Inventory: An Extension of the Davis Scoring Method." *Journal of Clinical Psychology*, IV (1948), 86-88.
- Muste, Myra J., and Sharpe, Doris F. "Some Influential Factors in the Determination of Aggressive Behavior in Preschool Children." *Child Development*, XVIII (1947), 11-28.
- Schettler, Clarence. "Objective Measurements of Personality Traits." *Journal of Personality*, XV (1947), 292-299.
- Siegel, Miriam G. "The Diagnostic and Prognostic Validity of the Rorschach Test in a Child Guidance Clinic." *American Journal of Orthopsychiatry*, XVIII (1948), 119-133.
- Staff, Personnel Research Section, the Adjutant General's Office. "The Army General Classification Test, with Special Reference to

- the Construction and Standardization of Forms 1a and 1b." *Journal of Educational Psychology*, XXXVIII (1947), 385-422.
- Surgent, Louis V. "The Use of Aptitude Tests in the Selection of Radio Tube Mounters." *Psychological Monographs*, No. 283 (1947), 40 pp.
- Timpany, Nancy. "Assesment for Foremanship." *British Journal of Psychology*, XXXVIII (1947), 23-28.
- Zubin, Joseph. "Recent Advances in Screening the Emotionally Handicapped." *Journal of Clinical Psychology*, IV (1947), 56-63.

THE CONTRIBUTORS

Donald E. Baier--Ph D., Princeton University, 1935. Interne in Clinical Psychology, New Jersey Dept of Institutions and Agencies, 1935-1936. Resident Psychologist, New Jersey State Hospital at Trenton, N. J., 1936-1937. Psychologist in State Mental Hygiene Clinic, 1937-1940. On active duty with the U. S. Army: Classification and Assignment Officer, 1941-1942; Personnel Procedures Officer and Asst. Chief of Section, 1942-1944, Officer-in-Charge, 1944-1946, Adjutant General's Office; Asst. Chief and Chief, Personnel Research Section, 1946- Lecturer in Personnel Psychology at N. Y. University and Dept of Agriculture Graduate School. Fellow, American Psychological Association Member, Psychometric Society, American Association for the Advancement of Science

Ray H. Bixler--M. A., Ohio State University, 1942. Psychologist, Akron Child Guidance Center, 1943-1944. Counselor, 1944, and Senior Counselor, Student Counseling Bureau, University of Minnesota, 1945- Author of articles in the *Journal of Clinical Psychology* and the *Journal of Consulting Psychology*. Member, American Psychological Association

(Mrs.) **Virginia H. Bixler** --M.A., Ohio State University, 1942. Rehabilitation Director, Summit, Stark and Mahoning County Tuberculosis Associations, Ohio, 1943-1944. Secretary, Case Work Consultant Committee Council of Social Agencies, Minneapolis, 1944-1945. Director, Vince A. Day Center, Minneapolis, 1945-. Author of articles in the *Journal of Clinical Psychology*. Member, American Psychological Association.

John M. Butler--B.S., University of Minnesota, 1939. Teaching Assistant, Instructor, and Administrative Fellow, University of Minnesota, 1939-1942. With the U. S. Army, Psychological Research and Classification Section, 1942-1946. With Advisement and Guidance Service, Veterans Administration Central Office, 1946. With Advisement and Guidance Division, Veterans Administration Branch Office, San Francisco, 1947. In the Dept. of Psychology, University of Chicago, 1947-. Member, American Psychological Association, Phi Delta Kappa, American College Personnel Association, Western Psychological Association.

A. A. Canfield--M.A., University of Southern California, 1947. Psychological Assistant, U. S. Army Air Force, for two years. Teaching Assistant and graduate student, University of Southern California, 1947-.

Kenneth E. Clark—Ph.D., Ohio State University, 1940. Instructor in Psychology, University of Minnesota, 1940-1942. Personnel Technician, Army Air Forces Technical Training Command, Atlantic City, N. J., and Personnel Research Section, Adjutant General's Office, 1942-1944. Assistant Classification Officer, U. S. Naval Training Center, Sampson, N. Y., 1944-1945. Staff, Enlisted Classification Section Bureau of Naval Personnel, 1945-1946. Assistant Professor of Psychology, University of Minnesota, 1946-. Member, American Psychological Association, American Association for the Advancement of Science, Menasha Psychological Association, Midwestern Psychological Association, Psi Chi, Phi Delta Kappa.

Doncaster G. Humm—Ph.D., University of Southern California, 1932; Sc.D., Bucknell University, 1945. Counselor, Sentous Junior High School, Los Angeles, 1921-32. Psychologist, Los Angeles Diagnostic Clinic for Neuro-Psychiatry and Psychology, 1931-1934. Lecturer, College of Medical Evangelists, 1929-31. Owner, Doncaster G. Humm Personnel Service. Author of articles on personality, temperament and clinical method. Co-author (with Guy W. Wadsworth, Jr.) of the *Humm-Wadsworth Temperament Scale*. Specialist in the application of clinical psychology to personnel work. Member, National Institute of Industrial Psychology (Great Britain), American Mathematical Society, American Statistical Association, Society for the Advancement of Management.

Philip H. Kriedt—M.A., University of Minnesota, 1947. Aviation Psychologist, Army Aviation Psychology Program, 1943-1945. Research Assistant, Industrial Relations Center, University of Minnesota, 1945-.

Ralph D. Norman—Ph.D., Ohio State University, 1946. Clinical Psychologist with N. Y. C. Children's court, 1936; N. Y. State Department of Correction, 1937-1938; Mississippi State Hospital, 1938; U. S. Public Health Service, 1939-1940. Graduate Assistant, Dept. of Psychology, Ohio State University, 1940-1941. With the U. S. Navy: Personnel Technician, Army Air Forces, 1942; Aviation Psychology and Classification Officer, 1942-1946. Psychologist, U.S.V.A., 1946-1947. Assistant Professor of Psychology and Assistant Director, Counseling Service, Princeton University, 1947-. Member, American Psychological Association, Eastern Psychological Association, New Jersey Psychological Association, Phi Delta Kappa, Alpha Psi Delta.

Margaret G. Powell—M.A., Ohio State University, 1947. Psychologist for Foster Care Bureau of Children's Services, Cleveland, Ohio, 1947-. Graduate Assistant on Staff of Dean of Women, Ohio State University, 1945-1946. Certified School Psychologist by State of Ohio. Member, Cleveland Psychological Association.

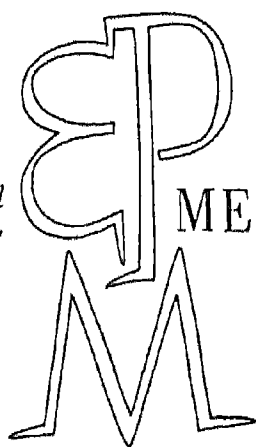
Arthur W. Sherman—M.A., Indiana University, 1942. Personnel Classification Officer and Aviation Psychologist, Army Air Forces,

1941-1946. Instructor in Psychology and Vocational Appraiser, Ohio University, Athens, Ohio, 1946-. Associate, American Psychological Association, Ohio Psychological Association. Member, American Association for the Advancement of Science, Midwestern Psychological Association, American Association of University Professors.

C. Harold Stone—M.A., University of Southern California, 1938. Candidate for Ph.D. degree, University of Minnesota. Teaching Fellow, University of Southern California, 1937-1938. Research Assistant, Committee on Educational Research, 1938-1940; Scholastic Counselor and Teaching Assistant, 1939-1940, University of Minnesota. Research Instructor, 1940-1941; Research Associate, 1941-1942, Employment Stabilization Research Institute. With the U. S. Navy: Staff, Training Division, Bureau of Naval Personnel, Washington, D. C. Assistant Enlisted Personnel Officer, Force Enlisted Personnel Officer, South Pacific Area and Force, Force Personnel Officer on Staff, Commander San Francisco Group. Social Science Research Council Fellow (Demobilization Award), 1946-1947. Author of articles on scientific film production, counseling and guidance, employee attitudes, job satisfaction and morale. Co-author of Local Labor Market Research (in press). Associate member, American Psychological Association, Minnesota Psychological Association. Member, National Vocational Guidance Association, Minneapolis Vocational Guidance Association, American Management Association, Society for the Advancement of Science.

Neil D. Warren—Ph.D., University of Southern California, 1934. With the Army Air Forces Psychological Program, 1942-1945. Associate Professor and Head of the Dept. of Psychology, Dean of Men, University of Southern California, at present.

EDUCATIONAL and
PSYCHOLOGICAL



MEASUREMENT

VOLUME EIGHT, NUMBER THREE, AUTUMN, 1948

<i>Prediction of Proficiency of Administrative Personnel from Personal-History Data.</i> J. P. GUILFORD and ANDREW L. COMREY	281
<i>Supervision of Counseling Services.</i> E. G. WILLIAMSON	297
<i>On Guttman's Scale Analysis.</i> ALLEN L. EDWARDS	313
<i>The Evaluation of Pairs of Tests for Guidance Use.</i> GEORGE K. BENNETT and JEROME E. DOPPELT	319
<i>A Study of Client Self-Selection of Tests in Vocational Counseling.</i> JULIUS SEEMAN	327
<i>A Note on Raven's Progressive Matrices Test.</i> H. J. A. RIMOLDI	347
<i>Individual Differences in Speed of Response in Mental Test Materials of Varying Degrees of Difficulty.</i> MERLE W. TATE	353
<i>A Comparison of Four Masculinity-Femininity Scales.</i> JOSEPH C. HESTON	375
<i>An Evaluation of Five Factors for Predicting the Success of Students Entering the New York State College of Agriculture.</i> J. P. HERTEL and FRANCIS J. DiVESTA	389
<i>A Meaningful Record of Tests.</i> EDWARD C. ROEBER	397
<i>The Contributors</i>	401

PREDICTION OF PROFICIENCY OF ADMINISTRATIVE PERSONNEL FROM PERSONAL HISTORY DATA

J. P. GUILFORD AND ANDREW L. COMREY

University of Southern California

ATTEMPTS to select administrative personnel by means of tests have not been as numerous or as successful as those in connection with other types of personnel. This can perhaps be attributed to the fact that administrative ability represents a varying complex of many different traits, few of which are easily isolated. The investigation reported here¹ represents an attempt to apply a biographical-data technique to the task of measuring administrative ability. It was believed that the biographical-data method had not previously been given an adequate trial in this particular field of personnel selection.

The encouragement to try this method stemmed largely from the success with which it was employed by the United States Army Air Forces (3, chapter 27) in the selection of pilots and other air-crew members. Although the resemblance between these tasks and those of administrative personnel is not great, both types of tasks are markedly complex. If the method could be applied to one complex task as a whole, it seemed reasonable that it might be successful for another.

Of particular interest in connection with the Army Air Forces studies was their investigation of the effect of directions upon the validity of the biographical-data test. It was found that standard directions with no mention of penalty for falsification gave results which were equally as valid as directions threatening severe punishment for falsifying the response. Directions which encouraged laxity in answering the questions gave dis-

¹ We are indebted to Dean Emery F. Olson and Professor John M. Pfiffner, of the School of Public Administration, who made this study possible by allotting funds from a research grant; also to Dr. Pfiffner for his continued interest and support; and to Dr. Paul E. Webb, E. C. Wills, Raymond F. Pollich, and Elizabeth Sands of the Los Angeles City Schools, who cooperated in the study.

tinctly lower validity. These results are important in any consideration of a biographical-data test for selecting employees, since the opportunity for falsification is great in this type of test.

Studies with life-insurance company employees have shown that biographical-data items are useful in predicting success in certain types of jobs. Several attempts have been made to utilize other printed tests to measure administrative ability with at least some degree of success. Thurstone (8) found that the Social Scale of the *Allport-Vernon Scale of Values* differentiated among certain federal government administrators. Mandell and Adkins (5) obtained validity coefficients greater than .60 for a top management group of executives on such tests as the linguistic section of the *American Council on Education Psychological Examination*, a *Civil Service Commission Current Events Test*, selected items from an *Interpretation of Data Test* of the Progressive Education Association, an administrative-judgment test, and an agency-organization-and-personnel test. Mandell (6) reported that the United States Civil Service had obtained promising results with the *Kuder Preference Record*. He believes that there is some evidence in favor of tests of mental ability and for interest inventories. Strong (7) presented some evidence to suggest that differences in interests between administrators and others may be expected. Achard and Clarke (1) reported definite indications that the measurement of interests can contribute to the selection of supervisory personnel. They used the *Vocational Interest Blank for Men (Revised)*, Form M, Scale CFS, with success and also the *Otis Self-Administering Test of Mental Ability, Higher Examination, Form A*. Uhrbrock and Richardson (9) found that mental-ability items were useful in differentiating between good and poor supervisors. A few biographical-data items were also helpful. These items related to age, schooling, confidence in blue-print reading ability, and military service. Mitchell (4) used biographical data in selecting sales managers. He found that better educated sales managers were more successful, but other items of biographical information did not prove to be effective.

Development of a Biographical-Data Inventory

As it has been used in the past, the familiar application blank represents an attempt to use biographical data, usually informally and unsystematically, in selection. Most of the research which has been done in various fields of personnel selection with such data has utilized an open-end type of question and a somewhat limited area of questioning. In the present investigation, the intent was to utilize a large number of questions covering a wide range of biographical information. In addition, it was decided that the questions should be put in multiple-choice form so that adequate statistical analysis of the results could be carried out. This feature of the test also provided for the use of standardized answer sheets and machine scoring. In this way much of the labor was eliminated, both for the subjects and the investigators.

The biographical-data booklet in its final form was a fourteen-page printed booklet containing 150 multiple-choice items. The items may be loosely classified into four different types: (1) childhood background and family life; (2) professional preparation; (3) health; (4) interests; and (5) early signs of leadership.

The items of the first group involved information concerning occupations, interests, and social activities of the parents, social habits as a child, and happiness of the early home environment. It was believed that the character of the home environment might do much to influence the development of those traits which are important in administrative ability. A sample item of this type was:

18. During most of the time before you were 16, you lived:
18—A With both parents.
18 —B With one parent.
18—C With a relative.
18—D With foster parents or non-relatives.
18—E In a home or institution.

The second group included items concerning the amount of education completed, types of subjects studied, scholarship and awards, major academic interests, and previous types of employment. Previous work with biographical data had indicated that amount of education was of value in the selection

of administrative personnel, so it was hoped that these as well as similar items might be significant. A sample item of this type was:

47. As a college student, you were:
 47--A A Phi Beta Kappa.
 47--B A Sigma Xi.
 47--C Graduated with honors. (But not 47--A or B)
 47--D A good student. (But not 47--A, B, or C)
 47--E An average student.

The third group of items concerned the health of the individual, both as a child and as an adult. The administrator is generally believed to be a person of excellent health and considerable energy. It seemed likely that individuals indicating a history of less than average health might prove to be poor administrators. A sample item of this type was:

- 25 Between the ages of 12 and 21, how often were you sufficiently ill to require hospitalization?
 25--A 0.
 25--B 1.
 25--C 2.
 25--D 3.
 25--E 4 or more times.

The fourth group of questions included a large number of items relating to interests, types of recreation enjoyed, reading habits, motion-picture and entertainment preferences, and social habits. In view of the suggestion of previous research that interests may be of value in the selection of administrators, it seemed important to include items of this type. A sample item was:

141. What type of radio program do you prefer?
 141--A Classical music.
 141--B Dance bands.
 141--C News commentators.
 141--D Plays.
 141--E Comedians.

The fifth, and final, group of items was devoted to questions concerning early signs of leadership. Since administrative work involves leadership in most cases, it was believed that

individuals who were leaders in their childhood and adolescence might later become good administrators. A sample item was:

35. How often were you a leader of your childhood "gang" activities up to the age of 12 years?
- 35 -A Always.
 - 35 -B Frequently.
 - 35 -C Occasionally.
 - 35 -D Seldom or never.
 - 35 -E (You were not a member of a group, or you can't remember).

Many other heterogeneous types of questions were included which cannot properly be placed in one of the four mentioned groups. For example, one item was:

101. Your own personality most resembles that of your:
- 101 -A Father.
 - 101 -B Mother.
 - 101 -C Brother.
 - 101 -D Sister.
 - 101 -E (None of these or you don't know).

Certain types of items were omitted because it was felt that they might prejudice the subjects against the research and hence do more harm than good. Questions pertaining to race, religion, and very personal matters were not included.

Item-Validation Procedures

The original sample included all of the regular principals and vice-principals of the Los Angeles City Schools. These fell into three groups, elementary, junior-high, and senior-high school principals. In preparation for the administration of the test, one of the investigators talked before most of the principals in the elementary- and junior-high-school groups to explain the purpose of the project, and to assure them that the information obtained would not be utilized by the school system for purposes of evaluating individual principals. It did not prove possible to offer this same type of orientation for the senior-high-school principals.

The biographical-data booklets were enclosed individually in an envelope together with an instruction sheet, a standard answer sheet, and an electrographic pencil. One of these

envelopes was then distributed via the official inter-school mailing system to each one of the principals to be filled out and returned. The principals were also instructed to return the booklets in sealed envelopes through the inter school mail. It would have been most desirable to have had the principals fill out their answer sheets in supervised groups. This kind of administration of the test was regarded as not feasible by the school authorities.

Returns from the senior-high principals numbered only 37, which constituted only a fraction of the total number in that group. For this reason, the results for that group were not analyzed. The elementary and junior-high groups were combined to be treated as one group. Of this combined group, 328 principals returned booklets, which constituted a return of 93 per cent. A few of those who did not return booklets were ill; the remainder either refused to turn them in, or ignored requests to do so. It was felt, however, that the remaining seven per cent of the cases not included in the analysis would not materially alter the results.

The 328 cases which constituted the final sample were subdivided into Groups A, B, and C. Group A was composed of 122 female principals selected at random; Group B included the remaining 123 females; Group C was composed of the 83 males included in the total elementary and junior-high sample. The female cases were separated into two groups so that a comparison could be made between the results of the analysis in one group and those in the other to obtain evidence as to whether any significant correlations obtained were not due to sampling errors. It was also planned to make a cross-validation study, i.e., derive a scoring key for valid items in one group and estimate the validity of the total score in the other group. In view of the item-analysis results, this step proved to be unnecessary. The men were put in one group because there were not enough of them to make up two groups.

The answer sheets filled out by the principals were divided into six groups, according to the division of the principals into high and low halves for each of the Groups, A, B, and C. Then, a scoring-machine tally was made of every response, for each alternative answer to every question for each of these six groups. For example, in question one, with five alternate answers, we

might find that 30 individuals in the high group and 10 individuals in the low group of Group A had marked the space corresponding to alternative 1-A. For Groups B and C, different tallies would occur.

With 150 questions, each with five alternative answers, a total of 750 categories were tallied for each of the six groups. Then, a correlation could be computed for each response for Groups A, B, and C separately. Thus, 750 correlations were possible for each of the three groups, or a total of 2250 correlations for the entire study. On the basis of visual inspection of the differential frequencies, plus previous calculation of the difference in frequency required for approximate significance, many of the correlations could be eliminated as being definitely below the level of significance. For the remaining ones, phi coefficients were computed, using an abac provided by Guilford (2).

The Criterion

One of the greatest difficulties in studies of the selection of administrative personnel is to decide who is and who is not a good administrator. This was perhaps the greatest difficulty and the weakest point in this research. The problem with which we were confronted was to divide each of the three Groups A, B, and C into two halves, higher and lower, with respect to administrative ability.

The only information available for this purpose consisted of the periodic ratings made by the school system for purposes of promoting principals to the next higher salary groups. This particular problem was studied for some months prior to the mailing of the booklets to see whether the ratings were of sufficient reliability to warrant a study. It was decided that the ratings exhibited sufficient *reliability* to enable us to carry out the research, but no determination could be made with respect to the *validity* of these ratings.

The ratings of each principal are generally made independently by three or more superintendents. There is a definite possibility that the basis upon which superintendents rate individual principals may be colored by something other than their administrative ability.

Many difficulties arose in connection with the evaluation of

these ratings. The elementary principals were rated in groups generally by means of a rank-order technique, in which each of approximately four superintendents arranged in rank order a stack of filing cards, each of which held the name of a principal being considered for promotion. The secondary or junior-high principals were generally rated on a five-category graphic rating scale with a spread of 100 points for each category. These categories included such traits as success in community relations, ability as a manager, success with personnel, and the like. A study of the intercorrelations between traits showed that a pronounced halo effect was operating, hence an average of the five different categories might just as well be used.

The difficulties with these ratings included the following ones, among others which we may have overlooked. The principals in Group A, for example, had not all been rated by the same superintendents, hence no satisfactory basis existed for comparing each principal with every other principal. Furthermore, some principals had been rated several times over a period of years while others had been rated only once, or not at all. Further, a comparison was made necessary between principals who had been rated on a graphic scale and those who had been rated by a rank-order method.

The problem of determining the reliability of these ratings proved even more of a dilemma. In order to estimate the reliability of the ratings, either different raters on the same individuals should be compared, or the same rater at different times, or both. These procedures require having data for all principals rated by the same superintendents on different occasions. Since only a few of the principals in a group had been rated in such a manner, it was very difficult to make a determination of the reliability.

In an attempt to get around these difficulties, the following simple procedure was used. A list of all the principals in each group was made. For every occasion on which he had been rated, the principal was given a relative position in comparison with the other principals rated along with him. That is, on the basis of an average of all his ratings, whether rank order or graphic rating scale, every principal in the particular group being rated at the moment was given a rank-order number.

This number was translated into a per cent position by the following formula:

$$\text{Per cent Position} = \frac{100(R - .5)}{N},$$

in which R is the principal's rank-order number and N is the number of individuals with whom he is being compared in that particular group. Each per cent position was then converted into a score from 0.0 to 10.0 in which 5.0 is approximately the mean. It was assumed that the individuals in each group were normally distributed with respect to the qualities rated and that the mean proficiencies of groups were actually much alike.

For those principals who had not been rated previously, special ratings were obtained and treated in the same manner described above. After all this manipulation of data, a list was available in which every principal had a numerical value beside his name which was roughly comparable with the numerical values of every other principal. In order to obtain one final representative score, an average was computed of all the scores for a principal, so that finally, one score for each principal was obtained.

In order to make an estimate of reliability, it was arbitrarily decided to include all principals who had been rated at least twice. Then, the first and last tabulated rating for each of these principals was taken. A phi coefficient (corrected for continuity) between these first and last ratings for 242 cases was .66, which was significant beyond the one per cent level.

In each of the three groups, A, B, and C, a median of the composite rating scores was determined. On the basis of these medians, the principals in each group were divided into higher and lower halves.

Results and Conclusions

Frequency distributions of the phi coefficients of significant size are given below. The levels of significance given were determined by reference to chi-square values, converted to corresponding phi coefficients. Although the uncorrected phi coefficients were used in determining their significance, the coeffi-

cients in this table, and below it, have been corrected for continuity in one variable ²

The 86 statistically significant responses listed in Table 1 were distributed among the three groups as follows: (1) 24 responses were significant in Group A only; (2) twelve responses were significant in Group B only; (3) thirty responses were significant in Group C only; (4) four responses were significant in the same direction in Groups B and C; (5) one response was significant in the same direction in Groups A and C; (6) two responses were significant in the opposite direction in Groups B and C; (7) one item was significant in the opposite direction in

TABLE 1
Distribution of Significant and Marginal Phi Coefficients Between Responses to Items and the Criterion of Administrative Proficiency

Phi	Group A	Group B	Group C	Total
.60-.64			1	1
.55-.59				0
.50-.54			1	1
.45-.49			1	1
.40-.44			2	2
.35-.39	2	1	4	7
.30-.34	6	2	10	22
.25-.29	16	8	"	31
.20-.24	8	8		16

Group A: $N = 122$; .05 level $z = 2.3$; .01 level $z = 2.9$

Group B: $N = 123$; .05 level $z = 2.3$; .01 level $z = 2.9$

Group C: $N = 83$; .05 level $z = 2.8$; .01 level $z = 3.5$

Groups B and C; (8) no response was significant in the same direction for all three groups; (9) no response was significant in the same direction for Groups A and B, the two female groups; and (10) a total of 57 items were involved in the 86 significant correlations for the three groups combined.

An analysis of the 57 significant items by groups reveals: (1) fourteen were significant in Group A only; (2) nine were significant in Group B only; (3) twenty were significant in Group C only; (4) five items were significant in Groups A and B. Of these five items, the alternative response results agreed substantially in one, disagreed in two, and neither

²This type of correction rests on the assumption that the criterion is actually continuous and that each response represents operationally a point distribution. The corrected coefficient should be numerically equivalent to a point-biserial r .

agreed nor disagreed in the other two; (5) five items were significant in both A and C groups. Of these, three were in agreement in the two groups, and two responses neither agreed nor disagreed; (6) three items were significant in Groups B and C. The responses in these items were significant in the same general direction in both groups; and (7) one item resulted in the same trend in all three groups.

Responses within items which reached the one per cent level of significance in Group A were: (1) participation in mild sports (golf, hiking, etc.) occasionally, $\phi = -.33$; (2) participation in mild sports (golf, hiking, etc.) seldom, $\phi = .35$; (3) participation in collecting (stamps, coins, antiques, etc.) frequently, $\phi = -.35$; (4) between the ages of 12 and 18, belonging to an organized group of children of own age without adult sponsorship, $\phi = .31$; (5) other than school work, reading during a large part of your free time between ages of 12 and 18, $\phi = -.31$; (6) participation in making speeches frequently, $\phi = .31$; (7) associating socially as a general rule with people younger than yourself, $\phi = -.29$.

For Group B, the following responses were significant at the one per cent level: (1) if the cost were the same, you would prefer to travel across country for pleasure by private automobile (rather than by bus, airplane, etc.), $\phi = -.38$; (2) as a child, you confided most in a brother or sister, $\phi = .31$; (3) you succeeded well in history as a college or school subject, $\phi = -.30$.

Responses reaching the one per cent level of significance in the male group were: (1) prior to the age of 21, you lived most of your life in a large city (over 500,000), $\phi = .46$; (2) you learned to swim at age 10, or below, $\phi = .35$; (3) between the ages of 12 and 18, you belonged to the boy scouts, $\phi = .40$; (4) you succeeded exceptionally well in history as a school or college subject, $\phi = .63$; (5) you succeeded exceptionally well in psychology as a school or college subject, $\phi = .35$; (6) participation in gardening occasionally, $\phi = .50$; (7) participation in travel frequently, $\phi = -.44$; (8) you are occasionally interested in making things, shop work (without necessarily having done so), $\phi = .38$; (9) you are fre.

quently interested in camp counselling, YMCA or YWCA work (without necessarily having participated), $\phi = -.38$, (10) your present weight is 170 to 189, $\phi = .37$.

From the eight items which gave significant results in the same general direction for at least two of the three groups, the following information is suggested: (1) the successful administrator's father generally employed more than five people; (2) confiding during childhood in the mother is not so auspicious as confiding in a brother or sister; (3) a child who when ill is put to bed, but with medication is more apt to be a successful administrator than if he had a physician called, had only home remedies, was merely kept at home, or had no special attention; (4) belonging to an organized group of children between the ages of 12 and 18 is worth while; (5) succeeding well in history as a school or college subject seems bad, but succeeding exceptionally well seems very good; (6) succeeding either well or exceptionally well in psychology as a school or college subject is not desirable for a female administrator, whereas succeeding exceptionally well in psychology is good for a male administrator, but only succeeding well is undesirable; (7) individuals who would like to have four or more children are not likely to be good administrators; and (8) those who were 40 to 50 years of age tended to be rated better as administrators, while those over 60 tended to receive poorer ratings.

In evaluating the significance of the results on an over-all basis, two approaches could be adopted: (1) a comparison of the number of significant *responses* in each group and in all three groups combined with the number of responses which would be expected to reach a standard of significance by random sampling among responses that actually correlate zero with the criterion; (2) a comparison of the number of *items* which contained at least one significant response with the number of items expected by chance, considering each group separately. Neither of these approaches is completely satisfactory, as will be pointed out.

The first approach reveals a total of 86 responses out of 2250 to correlate significantly at or beyond the five per cent level. Slightly more than 100 significant responses would be

expected if the normal sampling situation could be applied in this manner. Since the respondent is forced to choose only one of five alternative responses to an item, the 2250 judgments were by no means independent. For this reason, the first approach does not appear to be applicable. Because of the interdependence of responses, mostly in the form of *negative* correlation, a smaller number than 2250 should be used as the base; how much smaller is not known.

The second approach, evaluating the significance of the items as a whole, is probably the better of the two methods. If any response in a particular item demonstrates significance, that item as a whole may be considered significant. With a total of 150 independently answered items, approximately eight would be expected to yield correlations which equal or exceed the 5 per cent level of significance. Actually, 25 items were significant by this criterion in Group A, 18 in Group B, and 29 in Group C. This would suggest that a test composed of the significant items would have a predictive value significantly better than chance. While by this procedure there would seem to be a marked excess in the number of significant items for each group, there is at least one qualifying thought to detract from this conclusion. The items themselves are probably not independent. Just how the item intercorrelations may affect the number of items meeting the criterion of significance by chance effects alone is not clear. Since the item intercorrelations are probably positive (in the sense of the association of "good" features), chance factors might be expected to increase the dispersion of phi coefficients and thus to produce more than the normally expected number of significant items.

We are consequently forced to give attention to the comparison of item phi's obtained from different groups. There was not sufficient agreement among the results from the three groups with respect to the items which showed significance to warrant much optimism. Only eight of the 57 significant items showed agreement in at least two of the three groups.

The results of this study do not bear out previous findings with respect to the value of present interests for the measurement of administrative ability. There is a slight indication

that early environmental data may be sufficiently useful to warrant further investigation. The items used did appear to select male school administrators slightly more effectively than female administrators. This might indicate that female administrators in schools are a more homogeneous group than the males, or that biographical influences are not as uniformly effective with respect to administrative qualities in women.

A number of possible conditions might have accounted for our failure to obtain promising results. First, it might be that the items which were utilized were not sufficiently inclusive to reveal the areas of biographical data which would give the best results. Secondly, the school principals who participated might not have been truthful in giving their answers. It was mentioned earlier that one group had to be eliminated from the study because so few principals returned their booklets. If the other two groups merely returned the booklets without giving careful attention to their answers, the results would be invalidated. In view of the findings in the Army Air Forces study previously mentioned with respect to truthfulness of responses, lack of veracity does not appear to be a likely hypothesis, however.

A further consideration of importance is that of criterion validity. If the ratings upon which the division of the principals into higher and lower groups was made did not happen to represent real variations in administrative ability, it would not be surprising if negative results occurred.

The reliability of the criterion was sufficiently low to result in obscuring a definite relationship, if it happened to be one of low correlation. In such an instance it would be impossible to conclude that no real relationship exists.

Assuming that none of these considerations actually was responsible for the disappointing results obtained in this study, the problem of generalizing to other types of administrators and even to other groups of principals remains to be considered. In order to apply these results to other groups, we must assume that such administrators utilize the same types of traits for success in their work as the school principals, as rated, in this sample. This problem cannot be solved

except through further research with the biographical-data technique using other groups of administrators.

Realizing that so many variables are involved in the results reported here makes it difficult to draw any conclusions. A review of the results, however, definitely points to the conclusion that the biographical-data method has questionable value for the selection of school administrators. A reasonable presumption would be that this conclusion extends to similar information obtained in application forms or in interviews, or at least casts suspicion upon that information.

Summary

An attempt to develop a personal-history inventory for the selection of school administrators is described. A 150-question, multiple-choice type, inventory was given to more than 300 school principals and vice-principals. An item analysis was conducted in which responses to various items were correlated with success as a principal, measured by promotional ratings. Although a significant number of correlations was obtained separately in each of three groups of these principals, good agreement between the groups was obtained with only eight items. Possible explanations of the results were reviewed. It was concluded that the biographical-data method has only limited promise of usefulness for the selection of school administrators.

REFERENCES

1. Achard, F. H. and Clarke, Florence H. "You Can Measure the Probability of Success as a Supervisor." *Personnel*, XXI (1945), 353-373.
2. Guilford, J. P. "The Phi Coefficient and Chi Square as Indices of Item Validity." *Psychometrika*, VI (1941), 11-19.
3. Guilford, J. P. (Ed.) *Printed Classification Tests*. Army Air Forces Aviation Psychology Program Research Reports, Report No. 5. Washington, D. C.: Government Printing Office, 1947.
4. Mitchell, J. H. "An Experiment in the Selection of Sales Managers." *Occupational Psychology*, XII (1938), 308-318.
5. Mandell, M. M. and Adkins, D. C. "The Validity of Written Tests for the Selection of Administrative Personnel." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 293-312.

6. Mandell, M. "Testing for Administrative and Supervisory Positions." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, V (1945), 217-228.
7. Strong, E. K. Jr. "Interests of Public Administrators." *Public Personnel Review*, VI (1945), 166-173.
8. Thurstone, L. L. *A Factorial Study of Perception*. Chicago: University of Chicago Press, 1944.
9. Uhrbrock, R. S. and Richardson, M. W. "Item Analysis: The Basis for Constructing a Test for Forecasting Supervisory Ability." *Personnel Journal*, XII (1933), 141-154.

SUPERVISION OF COUNSELING SERVICES¹

E. G. WILLIAMSON
University of Minnesota

Major Objectives of Supervision

Improving Service There are three major objectives of the phase of administration known as supervision, two of which are more important than the third. One is the maintenance of continuously improving service to veterans, and I would stress the fact that it must go spirally upward, not remain static in that it reaches a certain level of effectiveness and then tapers off. Thinking back to the period following World War I we get perspective. At that time our knowledge of counseling was very rudimentary, and we would no more think of fixating at that level than we would think of permitting medicine to operate at the level of understanding of that day. The whole world has moved on in terms of a better understanding of the processes of effective counseling. I know there are some concepts of administration which fixate effectiveness at a particular level and maintain it. That is simply not possible in a world of human beings. If we had a world of machines, we could fixate at a particular level, but in the case of a professional service such as counseling we go backward or forward.

Developing Counselors. The second major objective is a little unusual for an administrator to be talking about. It is the continuous personal development of counselors themselves. This is an extremely important part of administration and much research is needed concerning the personnel problems of personnel workers. Why? Because the counselor is the instrumentality through which counseling is achieved. He is the only effective means by which counseling is achieved. All of the rituals are only half as important as the personality of the counselor. We keep forgetting this because adminis-

¹ This article is an adaptation of a talk given on January 9, 1948, in the Minneapolis Regional Office Advise ment Program, U. S. Veterans Administration

tration has been cast in the mold of a rigid structure, just as certain concepts of engineering are cast in a rigid mold. Counselors are constantly subject to environmental pressures, and sometimes these pressures interfere with the maintenance of a high level of effectiveness. The counselor himself is often as much in need of counseling as the client is, and any administrator that forgets it is not being a good administrator. It seems so axiomatic that it should be unnecessary to state it, yet in administration we almost refuse to do for the doctor what a doctor does for his patients. It is as though we counselors were a race apart without any problems. There is no other profession where there is a *perfect* fitting of worker and work, and we should not expect it of counseling, but we often do, and this is a part of supervision that is most seriously neglected. The old adage of "Physician, help thyself" is not often applied in the administration of a personnel program.

Controlling Quality.—The third objective of evaluation and supervision is of less importance, in my judgment. The so-called administrative quality control is less important as a direct object *per se*. We don't achieve it directly for one simple reason: counseling is a professional service, and quality in a professional service is best achieved by treating practitioners as professionals and by stimulating them and leading them to the desire to improve themselves. That is the way we get quality control, and not by setting up devices to see how many letters they have on their desks and all the other claptrap machinery by which one traps little people through administrative devices of one sort or another. An administrator has a hard time learning that the way to achieve quality is to give the professional plenty of elbow room, to let him use his imagination, and not to put him in a procedural strait jacket. If the counselor does not have imagination, nothing will substitute for the missing imagination, and it does little good to set up a manual of procedures. But we are still constantly trying to set up manuals of procedures to substitute for a lack of professional qualities, and without success. Given the professional capacity, procedures do not need to be written too specifically because the professional worker is able to adapt himself to the need of the client he faces.

To recapitulate, the three major objectives of supervision are: (1) the maintenance of continuously improving service to the veteran; (2) the continuous personal development of counselors themselves, because they are the instrumentality through which we achieve the first objective; (3) a quality control through treating counselors as professional workers. I will derive a number of generalizations from these objectives.

Methods of Supervision

Now we will turn to the methods of evaluation and supervision. I have tied up the two phases, evaluation and supervision, for particular reasons. If we are not dealing with professional service we do not need the concept of evaluation. All we need to do is to supervise the worker to keep him on the job. When we deal with the professional service, we need to emphasize evaluation because supervision has come to be a type of surreptitious searching for means to keep people's noses to the grindstone.

I shall discuss four methods of evaluation and supervision. I should preface my discussion by saying that I have in mind a particular kind of organizational setup, and some of these methods may not be particularly relevant to every situation. I am talking about general methods, and obviously many of them will have to be adapted to fit a particular situation.

Consultations. The first method may be described as non-regular consultations with a superior officer initiated by the counselor himself when he feels the need for aid. That is an ordinary state of relationship between professional workers. When a counselor feels something is out of his competence, he goes to his superior officer for aid through consultation. Now, the term consultation is not widely enough used in counseling circles. It is the conferring of two professionals about a situation that is not clear; it is the borrowing of ideas from equals, the exchange of relevant experiences. Please note that there is no implication of a loss of status on the part of one who seeks such a consultation. Moreover, there is no implication that he is disqualified for his job because he does not know all the things he ought to know. The technique of consultation is based upon the concept of professional ignorance which arises

out of the fact that one man specializes in what another man may not know, and that it is impossible for any one man to know all of the needed specialties. Therefore, these individuals of equal but different specialties confer and consult with each other to borrow services and to exchange ideas.

We ought to consult each other more frequently in counseling because we are in danger of developing the concept that a counselor should know everything when dealing with a particular client.

That should be all-sufficient. But common observation indicates that no matter how able the counselor is, he can never know all he ought to know to deal with any one client. Therefore we should abandon the idea of self-sufficiency in terms of technical background and understanding, and instead borrow the concept of consulting from medicine. But we have not developed yet to this sophisticated level of practice. If we can thus develop, we may consult with any superior officer about a particular case without having the superior officer think, "Why doesn't he know what he should know?"

Non-regular consultations with a supervisor whenever we feel the need for aid may result from uncertainty of what diagnosis we should place on a particular case. All counselors are uncertain at times, and all have puzzling cases because there actually are puzzling cases. This is a fundamental generalization that a counselor ought to keep clearly in mind. We ought to be impressed with the infinite capacity of the human animal to get into complex situations and difficulties. And our knowledge about his complexities fills only a little thimbleful in comparison with the huge reservoir of his latent capacity. Therefore, it is inevitable that many counseling cases are beyond our understanding. And if we do not recognize this truism, then there is something wrong with us as counselors. If we think we have all of the possible causes neatly packaged and labeled, then we are concealing the truth from ourselves. What counselor is not uncertain of many of his diagnoses?

Such a feeling of need for aid may also stem from a sense of uncertainty about the desirable outcomes of counseling for a particular individual. After all, in the way this world is

organized there is more than one road leading to Rome, and there is more than one thing that the client can do to get out of his current difficulties. Indeed there is often more than one occupational outlet for him. There is more than one type of training possible for him. There is more than one way to solve his maladjustment problem. I believe that any system of counseling which teaches counselors that there is a *sole standard approved* solution to any human problem is ridiculous and is based upon a simple-minded concept of human nature. Just as the capacity of the human being to get things fouled up is infinite, likewise his ways of getting them un-fouled are also infinite, and that is one of the beauties of the profession--it is an art in that there are many ways of helping the individual to solve his problems.

Parenthetically, we counselors and counselor-administrators need to be wary of trying to persuade people to follow *standard solutions*. We are always "solving" a client's problem with the dictum:- "the manual says" But we fail to perceive that manual writers in faroff places are always trying to confine reality to the narrow molds in which they sometimes think. If they do not think of techniques and solutions then, they are not important enough to think of. That seems to be the implicit assumption of a few manual writers. In contrast, I believe that there should be elbow room in counseling for the use of new ideas. No counselor should take away from his client the joy of inventing new ways of working out his own problems and no supervisor should deny such an inventive experience to any counselor.

If we apply such a principle to the supervision of counseling, then we see that a counselor may wish to consult his associates, because of some turmoil in answering a query or in giving uncertain advice to a client. The counselor may say to his superior, "That is the way this strikes me. What do you think about it? Will you check the information for me?" That is the kind of language that should be used between two professionals—one superior administratively, but equal professionally. Or the counselor may seek advice and aid because of a conflict of several possible actions or diagnoses, all of them equally plausible to him. Now, what counselor has not faced

the situation in which he can say: "It is either this, or this, or this? I can't see any difference between them." He knows very well that all possibilities cannot be right; therefore, there has to be a process of elimination, and there are days when neither the counselor nor the client can select from among several possibilities. Therefore, he needs consultation, the help of professional equals.

Case Reading.--The second method involves the careful reading of case histories, records and forms. The individual cases to be thus reviewed are selected and read by a designated staff assistant or superior official. The reading may cover a random selection of cases with regard to time or number, or type of case, or it may be a systematic selection of cases counseled on particular days of a week. I would rather favor the systematic selection in a rotation sense so that in a month's time the case reader would have covered every counselor and every type of case, such as rehabilitation involving particular kinds of disabilities, or a systematic review of types of occupational adjustment. But, on the other hand, there are many arguments for a random selection of cases. No system of evaluation and counseling can possibly cover every individual case handled by the agency. Yet I have seen some systems in which every case had to be handled twice, first by a counselor and second by a higher official. A professional system of counseling is not possible if a case has to be handled twice in such a manner. The implication of such a system is that the counselors cannot be trusted, and somebody has to put on the stamp of approval for each case handled.

Such a system of administration is based upon the concept of restricted and limited responsibility, a concept which is not applicable to counseling. Yet many of our current procedures in counseling are based upon a similar concept. But if we are dealing with professional people, we have to delegate to them lots of responsibility. We may have to spot-check cases for one of the three major objectives discussed above, but not for the reason of restricted and limited responsibility of the counselor. We have to keep in mind that there may be many levels of administrative responsibility, but there is only one level in professional counseling.

Case reading is a term used in social work. We pull out a

folder and we read it. We read it critically to see what was right and wrong in techniques used and to see what the counselor was trying to do; we make notes in the margin (unless our manual of procedures forbids that!) or we dictate a criticism of procedures used; or we consult with the original counselor to see what he had in mind, but failed to describe clearly in his case notes. We do not call him on the carpet - we simply try to achieve an understanding. That is the whole purpose of case reading. We simply want to say, "Tell me what you had in mind. It is not clear from this record." The language we use is the language of professional equals, not the language of a superior officer even though the administrative responsibility is clearly present and defined. Or we may, instead of making notations in the margin of the record, have conferences informally, or use one of the other methods discussed below.

Case Conferences. The third method of supervision is the case conferences, weekly or bi-weekly. An oral presentation of a case is made by a counselor. It may be a case selected at random; it may be a case brought in by the registration officer; it may be the counselor's most perplexing case of the week or his best case of the year. Assuming that counselors have at least one "best" case each year, it is good supervision to let him get the feeling of pride that comes with discussing with his associates a case well done. When the case conference is concluded he may not be so proud of it, but he should have the satisfaction of saying, "This is my best case." The essential feature of this case conference is that people pool their ignorance publicly. They simply sit around a board as a group of professional equals and discuss and criticize each other's professional work, always with the polite language of professionals, but being quite critical of others on the assumption that there is more than one technique to be learned because the same set of case data may be validly interpreted in at least two different ways by two different sets of counselors. It is a very healthy thing for each counselor to hear other persons interpret the counselor's own data. Of course, the presiding counselor must establish, and maintain impartially, "ground rules" for this scrimmaging, otherwise one individual will lecture his colleagues on what is "the revealed truth!"

We often rotate counselors in this presiding chair so that

every counselor sits in the seat of honor from which vantage point he discusses his own case and gives his own understanding of the nature of counseling. This situation probably is one of the most effective techniques of all in supervision because in effect we are permitting each counselor to be the evaluator and supervisor of every other counselor. There is a certain ego status outcome of this experience that is very desirable because we do not lose status when professional equals criticize equals as much as when an administrative superior criticizes, even if he is one of the professional counselors.

Evaluation by an "Outsider". -The fourth method of evaluation involves non-regular inspections by visiting consultants who, after reading case records, may make an oral or written critique of each case read, or may participate in a case conference and even make a written critique after they hear oral presentations by counselors. It is a good thing to bring in to the inner circle of counselors someone from the outside because one of the characteristics of human beings is that they settle down into a comfortable rut of procedures very readily. We all like to regularize and habituate our lives, even our professional lives. This characteristic sometimes is one of the most discouraging outcomes of dealing with human beings, even with professionals, because of the difficulty of climbing out of comfortable ruts. People like habits so much that they do not want to change and thus we hear some counselors say, "Well, I've counseled this way nigh onto forty years; it's been good enough so far and I don't see why I should adopt these new fangled ideas, especially those non-directive techniques!" Whenever we hear counselors say such things, counseling is going rapidly downhill. To offset this tendency to settle into comfortable ruts we sometimes must prod people into progress. This generalization is as much true of professional counselors as it is of politicians, school teachers, parents, even students, and certainly it is true of administrators!

Viewing counselors as people with their own peculiar psychology, a most interesting phenomenon of the past decade is their psychological reactions to the prodding of the School of Reflectors. Much heat and a little light have been generated in resistance to Rogers' ideas, but his followers have prodded us

into thinking critically of some things we have complacently taken for granted. We thought we had collected all of the carved tablets from Mount Sinai but now we find that there are a few we had overlooked.

Yes, we learn much in conferences with *Auslander* even though the experience is often uncomfortable. When we say that we have counseled in a particular way for forty years, then that is the time when an administrator ought to be the most worried. Indeed an administrator ought to worry most about his own tendency to regularize and routinize procedures. Oftentimes he has the largest amount of resistance to new ideas. He smugly says too often: "We settled that problem five years ago. Why do you reopen it?" Such a type of mental makeup in counselors and administrators and in manual writers as well, is just holding back progress. The judicious use of the "visiting fireman" technique upsets such smugness and a good administrator always chuckles to himself whenever his staff is upset following a visit from such an outsider. He thinks: "I have this staff stirred up now. Maybe that experience will give them some new ideas." But the counselors and the administrator himself will inevitably settle down again. Therefore a good administrator is one who knows when to prod himself and others and when to rest before he prods them again. Of course he prods in ways which are not harmful to mental hygiene although some administrators prod too much, I suppose, because they have an overactive thyroid or because they are building an empire too rapidly. On another occasion we should explore the personnel problems of personnel administrators.

Criteria for Evaluation

What are the criteria for evaluation? What is it we are looking for when we read a case as supervisors? We must have something in mind, some guideposts, some guiding principles by which we evaluate. What is the difference between two cases, between two counselors, etc? How do we differentiate between "good" and "bad" counseling? My experience yields six criteria but there may well be others more important growing out of the experiences of other counselors and administrators.

Congruence of Objectives. The congruence between the expressed desires of the client and the counselor's advice is one of the simplest criteria. The counselee may say, "I have this objective or action in mind. I have come to see whether it is feasible." What happened afterwards in the counseling situation? The veteran came in with one mental set; he went out with another. What is the explanation for the change? If a counselor does not reveal why there is a discrepancy, he may not be a good counselor. If he is not aware of the discrepancies, he has missed a significant point. What is the explanation and is it a valid explanation? It may be that the counselor has the case neatly ticketed with a particular vocational objective in mind. But when the *total* case record is reviewed—background, schooling, desire to establish a home today or at an early date, amount of supplementary resources, aptitude, etc.—the projected vocational objective is not feasible. Therefore, the counselor and the client searched for a satisfactory substitute objective. That kind of an explanation of the putting together in a congruent and balanced "whole" of all the known pieces of relevant data is evidence of good counseling.

Counselee Satisfaction. Another bit of evidence of good counseling is the quality of the client's transition from one stage of insight to another. When the counselee comes in with one mental attitude and leaves with another, the transition usually has to be made without an emotion disruption. Otherwise we have a dissatisfied client, and dissatisfied clients are not usually the product of "good" counseling, although this bit of evidence, like all others, must always be interpreted in the context of the total case record.

There are certain kinds of dissatisfaction that are inevitable, but we have to ameliorate a feeling of dissatisfaction, otherwise you have a client who says, "The counselor told me I *had* to do such and such." After all, this is a social democracy and clients should not be *told* that they *have* to do "such and such."

This criterion is very important—satisfaction of the client with the service and with the advice. Satisfied clients are a very important end in themselves, and there is a very real reason why they are. Satisfied people often tend to think more clearly because they are not distracted by frustrations.

They tend to think more clearly; they tend to exploit the possibilities of the occupation if they feel it is one that is congenial to them emotionally. For this and other reasons, evidences of dissatisfaction found in the case record must be explained by the counselor.

I have made a side reference to the obvious fact that certain dissatisfactions are inevitable and that some can be ameliorated but not prevented. That is true, for example, of an individual who has made an irrational vocational choice—a person with an IQ of 110 who wishes to be a doctor because of a fixation on the family physician—and is thus doomed to a certain kind of dissatisfaction. That is, it is inevitable because of the way our occupational world is organized. Such a type of dissatisfaction is not what I am talking about. Good counseling is possible whether or not the counselor could ameliorate such a type of dissatisfaction. But he does not do it by saying, "You don't have the right amount of IQ." The counselor can handle that type of case effectively even if he does not convince the boy emotionally, as well as intellectually, that a substitute occupation is a better one for a person of his capabilities.

You know well the techniques by which this result of counseling is accomplished. They are not as much intellectual as emotional in character. Sometimes an emotional transference is the only way it can be done. The boy attaches himself emotionally to the counselor and will accept the counselor's advice simply because he likes the counselor. Therefore rapport is a very important technique in achieving some legitimate outcomes of counseling. It helps to gloss over disappointment, merely because the counselor has a sympathetic insight and shows it in the way he treats the client. Appreciation of and insight into emotional subtleties are important in counseling. These are often the signs of "good" counseling the supervisor is looking for as he reads the typed and written record.

Counselee Understanding.—Indications of the adequacy of understanding by the client of the information and especially of "advice" given by the counselor are important. One of the most difficult puzzles a counselor must solve is to get the client to see what it is he is being advised to do. The distortions of understanding by the client are often as ingenious as those

recorded by students in class examinations. The inability of people to hear accurately is an alarming thing. Have you ever heard a client reproduce the advice you gave him? It does not sound like what you thought you had said originally. These distortions are not intellectual in their origin—there are all sorts of psychological factors involved. But it is not only understanding of advice but, more important, the basis for that advice that is involved in good counseling. In a social democracy one of the end goals is to help people to learn to stand on their own feet in regulating their lives. That is why counseling is so basic in our culture. If that goal is to be achieved, then the client must learn the basis of advice given to him and this means that he must learn methods of solving personal adjustment problems. Hence the supervisor of counseling reads the record for evidences of effective teaching of methodology—not *merely* for an effective solution of the client's current adjustment problems.

Last year when I was in Germany I observed a different type of counseling. I found an "assignment" counseling. One vocational counselor argued that in Germany the counselor *never* insisted that an individual had to follow an occupational choice, and yet at the same time the counselor reported that the parent had the final word on what occupations should be chosen by a son or daughter. I asked, "Does that mean the parent dictates?" The reply was a return to the statement that the student was free to choose—a curious contradiction without insight. But during the Nazi regime a person had no freedom of vocational choice. That is, unless he "chose" the occupation the "state" thought most necessary for the state. That is a widespread concept on the continent, that the state does the choosing in subtle or indirect ways. Such a concept has a familiar ring because counselors in military establishments "assigned" persons to particular jobs and establishments. But this method is not what we use if we desire people to learn to solve their problems.

In effect we are trying to help our counselees to be their own counselors. This need not alarm the administrator that he will run himself out of business. Rather, the more intelligent people get about counseling, the more they come to experts.

Most persons will be amateurs in counseling themselves. Even the experts, when it comes to their own problems, are amateurs. If you are a counselor and have children, you know you are often disqualified and unqualified to counsel your own children and you must "borrow" services from another expert. There is something about the nature of counseling relationships that make this situation almost inevitable.

Increased self-understanding as a preparation for later readjustment- I wish I had more time to give to it because it is the most important outcome of counseling, the one least understood, and the one on which we have the fewest established facts and techniques. How do you teach people self-understanding? It is a very difficult problem, and just "telling" people what you advise and why you advise it is not often an effective device. There has been enough experimentation to indicate that just "telling" our client why we advise something, does not necessarily give him deep understanding. But the case reader searches for bits of evidence of the counselor's attempts to teach self-understanding and methodology of achieving self-understanding.

Occupational Opportunities.—Congruity between vocational advice and the training opportunities available, or those which could be made available, is significant. I remember a case that illustrates this importance in an unusual manner. It happened about 1932 during the depression. A girl came to me for advice, and said she wanted to go into teaching. We collected case data and intelligence tests and found everything was quite favorable, so I advised her to go ahead. You will remember the surplus of teachers in those days. I was telling a visitor about this case and she said, "I hope you didn't encourage her to go on. There are thousands of teachers who cannot get a job today, and you should not encourage a student to go into teaching. How do you justify your advice?" I replied: "I explain it by the fact that this was an unusual individual. There are some teachers being hired, and I assume that a well-qualified person is not going to have a hard time getting a job, even in this depression period. There is always some occupational mobility, a few persons are being hired even today. Secondly, this individual is *beginning* her professional

training, and can you predict what employment conditions are going to be four or five years hence? These two reasons justify a little caution in using present day labor market conditions as the *sole* basis of vocational guidance of college *freshmen*."

All this is by way of warning counselors to be aware of the broad context of counseling data. They should not tie vocational guidance too closely to immediate employment opportunities. Training usually takes a little time, and employment opportunities have a way of fluctuating. I know of no basis on which we can predict, with a great deal of certainty, next year's employment opportunities, and certainly today's employment opportunities are not necessarily going to be representative of those available three or four years from now. Caution should be exercised in interpreting employment opportunities in vocational advisement. And the case reader searches for indications that both the counselor and the counselee understand such types of cautions and safeguards in interpreting and using case data.

Personality Perceptions.—The fifth criterion is the adequacy of consideration given by the counselor to personality factors. Some counselors seem to feel that if the individual has basic aptitude, that is all that is necessary for "success." They seem to have forgotten that most occupations are *social* occupations demanding basic skills in working with other people, that social adjustment in itself requires basic aptitudes. We, therefore, have to give careful consideration to the personality makeup of the individual, *even in vocational guidance*. Contrary to popular thought, mental hygiene is not restricted to explosive emotional troubles, rather is it a normal part of this particular kind of civilization we have today. Therefore, personality counseling is a very important generic technique, and the counseling supervisor is alert to the counselor's perceptions of this phase of the counselee's adjustments.

Adequacy of Records.—The sixth and last criterion is a very routine, mundane thing, and yet it is a rather important mark of a good counselor—the adequacy of notes and records. I am a believer in making standard interview forms and then throwing them away and giving the counselor a blank sheet of paper on which to record what he wishes to have recorded

in the history. I know of no way so to standardize counseling that you can embalm its substance all in a narrow box on a printed form. In sharp contrast, the free-flowing type of record-keeping gives the counselor elbow room to express what he seems to think is important about a particular case. Rigid procedures and tight fitting forms grow from attempts to standardize counselor and counseling. And I am persuaded that neither can, or should, be fully standardized or even fully structured. Every professional service encompasses some elements of imaginative inventions and counseling is no exception.

Anyone who re-reads another's case notes must be able to reconstruct the entire case situation, and we cannot do that from a form. We can better approximate a full understanding of counseling processes from a free running account or description. There is some indication that a fully dictated account is more adequate in the sense of being a more valid report of what actually transpired than is a brief summary of the outcome of the counseling. However that may be, we cannot evaluate counseling until we can reconstruct the processes that lead up to the outcomes. If we desire the outcomes of counseling to be recorded in a neat form with a few phrases, and if I review and evaluate and supervise, then I want to be able to reconstruct the processes by which the counselor thought he and the client arrived at that end product. If we merely evaluate in terms of end product, we cannot evaluate counseling adequately. In order to understand it is necessary to have more complete notes than are usually afforded by standard forms.

I have discussed briefly three objectives of supervision in the field of counseling from the viewpoint of an administrator charged with responsibility for maintenance of an effective personnel program for clients. Four methods of achieving these objectives were outlined from the viewpoint of a supervisor of counseling in a public agency devoted to a program of personal services to clients. Finally, six criteria were described in terms of their contribution to an evaluation of the effectiveness of counseling services.

ON GUTTMAN'S SCALE ANALYSIS¹

ALLEN L. EDWARDS

University of Washington

A RECENT article in this journal by Louis Guttman (6) illustrates several of the difficulties with scale analysis at its present stage of development. One of these difficulties concerns the location of cutting points for various items. With perfect reproducibility there would be, of course, no problem. The cutting points for an item would simply be the dividing points in the rank-order scores where the response shifts from the more to the less favorable category. Items are perfectly reproducible when from rank-order scores we can reproduce all the responses to the individual items. All subjects with rank-order scores above the cutting point would respond by checking the more favorable alternative and all those below would respond by checking the less favorable alternative. There would be, in the case of perfect reproducibility, no overlap in the responses of the subjects below and above the cutting point. But perfect reproducibility is an ideal case and is not likely to be encountered in practice. The question thus arises as to how the location of cutting points for items which are not perfectly reproducible is to be determined.

Guttman offers the suggestion that the cutting point should be located so as to minimize error (6, p. 258). But since it is also true that the reproducibility of an item can never be less than the frequency in the modal category, we have another rule that "no category should have more error in it than non-error" (6, p. 261). Now these might seem to be obvious and easy rules to follow, but apparently Clark and Kreidt (1) experienced some difficulty in agreeing in the treatment of items when working independently according to the rules. In terms of the data presented in Guttman's latest article (6), it would seem that there is some inconsistency there also.

¹ I would like to express my indebtedness to my colleague, Professor Paul Horst, for the many discussions we have had concerning scale theory, and for a critical reading of this paper.

Guttman finds a coefficient of reproducibility of 89 per cent for the 7-item test concerning attitude toward *A Nation of Nations* (6, Table 3, p. 259) and this area is considered scalable.² All of these questions meet the rule concerning less-error than non-error in the categories. However, for the 7 items, all dichotomous scoring, concerning attitude toward *Black Metropolis* (6, Table 7, p. 275), Guttman arrives at a different conclusion. "Several of the questions have more error than non-error," he states (6, p. 272). The content of this area is therefore judged non-scalable.

If the reader will re-examine Guttman's Table 7, however, it can be demonstrated that Guttman's conclusion does not necessarily follow. The cutting points *can* be placed so that *no category has more error than non-error* and the coefficient of reproducibility will be found to be 88.3 per cent. The cutting points which yield this coefficient can most easily be given, for the reader who wishes them, by giving the dividing points in the frequencies at the left of Guttman's Table 7 for the various items. Thus, for the first item, the cutting point falls between the 9th and 10th individuals, counting from the bottom up. The location of the cutting points for the other items are as follows: (Item 2) between the 11th and 12th individuals; (Item 3) between the 13th and 14th individuals; (Item 4) between the 16th and 17th individuals; (Item 5) between the 5th and 6th individuals; (Item 6) between the 19th and 20th individuals; (Item 7) between the 7th and 8th individuals.

If the cutting points given above are used, no item will have more error than non-error, if error is interpreted as Guttman apparently does with respect to the 7 items concerning *A Nation of Nations* which gave a coefficient of reproducibility of 89 per cent. Furthermore, we have here only 41 errors compared to the 40 errors which Guttman counted for the *A Nation of Nations* test. Why, then, is one area judged scalable and the other not?³

² There are other criteria for scalability as outlined by Guttman in various publications (4, 5). We are not here concerned with scalability as such, but only with the computation and interpretation of the coefficient of reproducibility.

³ That other criteria are to be applied in judging scalability has been pointed out before. Guttman, however, does not mention these as the basis for his rejection of the hypothesis of scalability of the items under consideration. The only reason given is that "too much error" is still present (6, p. 272).

The results of the analysis cited, however, are definitely misleading. And this statement applies equally well to Guttman's analysis of the *A Nation of Nations* test. The reason is that some of the cutting points used in counting errors fall *within* a given score interval. If cutting points are to be rigidly defined, then they must fall between scores. Only in this way will the coefficient of reproducibility have meaning. It is Guttman's contention, for example, that the coefficient of reproducibility measures the degree to which it is possible to reproduce item responses from rank-order scores. This is true, if the coefficient of reproducibility is computed in the manner to be described, where cutting points fall between scores. It is not true if the coefficient is computed as Guttman does for his Table 3. For the data given there, Guttman believes that it is possible to reproduce from a person's rank-order score his response to each question with 40 errors for the entire sample (6, p. 260). Let us examine this claim in some detail.

Tabulating the most frequent patterns of response from Guttman's Table 3, indicating the weights and the frequencies of each score, we arrive at our Table 1.

Now if we predicted the response patterns shown for the scores at the left of Table 1, we would make the errors given at the right of the table. Thus 40 of the responses, for the sample at hand, would be incorrectly predicted. But there is an obvious difficulty with some of these patterns: those marked with an X are inconsistent with the notion of a scale. For we have, for each of the patterns marked with an X, subjects with lower rank-order scores giving more favorable responses to items than subjects with higher rank-order scores, or subjects with higher rank-order scores giving less favorable responses to items than subjects with lower rank-order scores.

The logical procedure, and one consistent with scale theory, would seem to indicate scoring the individuals with "inconsistent" patterns as the nearest "scale type" (4). That is, if we make the patterns for these scores consistent with the rest of the data and with the notion of a scale, we would have the response patterns given in Table 2. We see that scores 10, 8, and 7 would now be scored with scale type 9; score 5 would be scored as scale type 4; and score 1 would be scored as scale

type 0. The pattern for score 12 must also be revised. Essentially, what we have done is to treat the set of items at hand *as if* they were perfectly reproducible for the sample. If this were true, then the scores 10, 8, 7, 5, and 1 would not appear in the sample, nor would the patterns of response giving rise to these scores; the response pattern for score 12 would also be revised as indicated.

TABLE 1
Predicted Patterns of Response for Scores on a 7 Item Test and Resulting Errors of Prediction

Scores	Pattern of Response to Items	Errors
14	2 2 2 2 2 2 2	0
13	2 2 2 1 2 2 2	0
12	X 2 2 2 2 2 0 2	3
11	1 2 2 1 2 2 1	2
10	X 1 2 2 1 0 2 2	2
9	1 2 2 1 0 2 1	7
8	X 1 2 2 1 0 2 0	0
7	X 1 0 2 1 0 2 1	5
6	1 2 2 1 0 0 0	8
5	X 0 0 2 1 2 0 0	0
4	1 0 2 1 0 0 0	5
3	0 0 2 1 0 0 0	6
2	0 0 2 0 0 0 0	0
1	X 0 0 0 1 0 0 0	2
0	0 0 0 0 0 0 0	0

TABLE 2
Revised Patterns of Response and Resulting Errors of Prediction

Original Scores	Pattern of Response to Items	Revised Scores	Errors
12	1 2 2 1 2 2 2	12	5
10	1 2 2 1 0 2 1	9	4
8	1 2 2 1 0 2 1	9	1
7	1 2 2 1 0 2 1	9	6
5	1 0 2 1 0 0 0	4	2
1	0 0 0 0 0 0 0	0	3

If the statements above are true in the case of perfect reproducibility, and if we now compute the coefficient of reproducibility based upon deviations from the patterns that would appear in the case of perfect reproducibility, we have a legitimate measure of the degree to which responses to the items can be reproduced from rank-order scores. This is an objective and reliable definition of the coefficient of reproducibility. Our predictions of item responses are based upon the assumption of a perfect scale or perfect reproducibility for the

sample at hand; the coefficient will indicate the degree to which we approach 100 per cent reproducibility.

We shall have the same number of errors for all scores except the ones for which the response patterns have been changed to accord with the notion of a scale. The errors we originally had, 40, have now been increased by 9. This gives us a total of 49 errors or an increase of 22.5 per cent over the original 40 counted by Guttman. The coefficient of reproducibility is changed relatively little, of course, by this increase in this particular instance. It is now 86 per cent compared to Guttman's original 89 per cent. With a larger number of cases and additional items, the difference might be quite significant.

There is still another point to be mentioned concerning the reproducibility coefficient for a given set of data. For the sample at hand, Guttman's Table 3, it is obvious that subjects with scores of 14 and 0 cannot possibly contribute any error. There is one and only one way in which these scores can be obtained and the response pattern for subjects with either of these scores will be perfectly reproducible regardless of any other factors. In the present sample there are 6 such cases contributing a total of 6x7 or 42 responses out of the entire sample of 350 upon which the coefficient of reproducibility is based. Obviously, with an excess of cases at either extreme or both extremes of the scale, reproducibility might seem much better than it actually is. For example, if we eliminate these cases, and use only the data which permit error, we have only 84 per cent reproducibility.

The criticisms raised here are not directed at the theory of scale analysis. Guttman has made a signal contribution to test construction—a contribution that is not to be disparaged. There is, however, a definite need, at the present time, for a clarification of the problem of locating cutting points and computing coefficients of reproducibility, so that the technique has a higher degree of objectivity in the hands of various investigators. The suggestion offered here is that we, for a given sample, assume perfect reproducibility, and make our predictions of item responses on this assumption. Error may then be measured in terms of the number of responses departing from the pattern predicted. Cutting points for items would thus be

rigorously defined and would always occur between ranks. Scores which are inconsistent with the assumption of perfect reproducibility would be scored as the nearest scale type consistent with the notion of a scale. This method should lead to essentially the same results when used by various investigators and has been advocated by Goodenough (3).

The difficulties mentioned above are relatively minor ones. The major problem, which Guttman has not as yet solved, is the original selection of the set of items to be tested for scalability. Techniques of item analysis, in the sense in which this term is ordinarily used by test constructors, are apparently ignored by Guttman in favor of an a priori selection of items by the investigator. Out of the hundreds of items which could be written concerning any defined area of content, Guttman selects but 10 to 12 to test, *as a set*, for scalability. Why one particular set is selected rather than another remains something of a mystery, if we may take the sets presented in various papers as representative of the a priori approach advocated. To say that we should look for items that seem to be homogeneous in content (5, p. 461) adds relatively little to what many test constructors have known for some time.

Scale analysis, it may be said, permits us to evaluate how well we have selected a set of items with respect to homogeneity *after* they have been selected. It does not tell us how to select items which will be homogeneous.

REFERENCES

1. Clark, K. E. and Kreidt, P. H. *An Application of Guttman's New Scaling Techniques to an Attitude Questionnaire*. In press.
2. Edwards, A. I. and Kilpatrick, F. P. *Scale Analysis and the Measurement of Social Attitudes*. In press.
3. Goodenough, W. H. "A Technique for Scale Analysis." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT* IV (1944), 179-190.
4. Guttman, I. *Questions and Answers about Scale Analysis*. Research Branch, Information and Education Division, Army Service Forces, Report D-2, 1945.
5. Guttman, I. "On Festinger's Evaluation of Scale Analysis." *Psychological Bulletin* XLIV (1947), 451-465.
6. Guttman, I. "The Cornell Technique for Scale and Intensity Analysis." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VII (1947), 247-280.

THE EVALUATION OF PAIRS OF TESTS FOR GUIDANCE USE

GEORGE K. BENNETT AND JEROME E. DOPPELT

The Psychological Corporation

In the process of educational or vocational guidance, it is the usual practice to employ several tests which are administered to the client for the purpose of revealing his strengths and weaknesses so that he may decide upon a career which will utilize his superior capacities and avoid, so far as possible, dependence upon those characteristics in which he is least able. Consequently, it is desirable to have tests which do not correlate very highly with each other, thereby measuring relatively independent traits. Another characteristic of importance in the selection of tests for this purpose is high reliability so that differences in scores shall be, so far as possible, true differences rather than accidental ones.

Whenever two tests are considered for use in combination, it is convenient to have some means of determining their differential power. Some investigators have used the corrected coefficient of intercorrelation or a function of the coefficient of correlation, the coefficient of alienation (k), as a means of estimating the degree of independence of a pair of tests. Neither of these statistics lends itself readily to meaningful interpretation in terms of differential power. Some years ago T. L. Kelley¹ proposed another method which he called "the proportion of differences in excess of the chance proportion." This method also depends upon the reliability of the tests and their intercorrelation. It has the advantage, however, of being directly meaningful since it estimates the proportion of all differences for that pair of tests which will represent true differences in ability.

¹ Kelley, T. L. "A New Method for Determining the Significance of Differences in Intelligence and Achievement Test Scores." *Journal of Educational Psychology*, XIV (1923), 321-333.

Although the method for computing this statistic was described so long ago, instances of its application are very rare. David Segel seems to have been the only experimenter to publish a report of its use.²

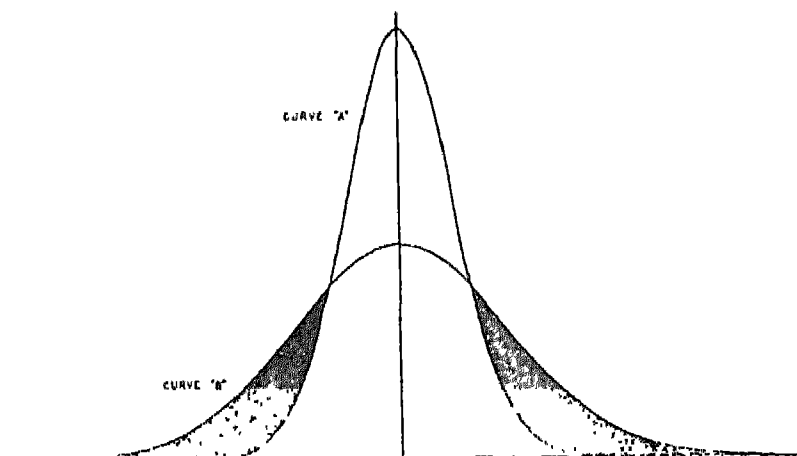


FIG. I Graphic illustration of the "proportion of differences in excess of the chance proportion" "A" and "B" are normal curves of unit area with $\sigma_B = 2\sigma_A$. Curve A represents the distribution of differences resulting solely from the unreliability of the tests. Curve B represents the distribution of obtained differences. The shaded areas under curve B represent the proportion of differences in excess of chance. For example, this condition would result from two tests with respective reliabilities of .80 and .90 (mean $r_{11} = .85$) where the intercorrelation is .40. The proportion of differences in excess of chance is .32.

FORMULAS. Standard deviation of differences due to unreliability of tests (Curve "A" above)

$$\sigma_d \cdot \alpha\omega = \sqrt{2 - r_{11} - r_{22}}$$

Standard deviation of obtained distribution of differences (Curve "B" above)

$$\sigma_d = \sqrt{2 - 2r_{12}}$$

Further discussion of differences pertaining to individuals may be found in Kelley, T. L. *Fundamentals of Statistics* Harvard Univ Press, 1947, pp. 413-419

Figure I will aid in understanding the Kelley method. Two tests are administered to each member of a group. The scores are converted to standard scores and the difference is obtained for each individual. The distribution of such differences is represented by Curve "B" in Figure I. It is evident that the width of this distribution is inversely related to the intercorrela-

² Segel, David. *Differential Diagnosis*. Baltimore: Warwick and York, 1934.

tion of the tests. However, since the tests are not perfectly reliable, some differences will result solely from errors of measurement. This hypothetical distribution of differences resulting from the unreliability of the tests is shown by Curve "A" in the same figure. The area under Curve "B" and outside Curve "A" represents the "proportion of differences in excess of the chance proportion."

To compute this index, it is necessary to find the values for the two standard deviations, the ratio of the distribution of errors to the total distribution, and finally, to consult a table prepared by Kelley which gives the area unique to Curve "B" for selected values of this ratio.

In connection with the development of a series of differential aptitude tests, there was frequent occasion to perform this computation and, as a labor-saving device, the nomograph appearing in Figure II was prepared. Values determined from the nomograph correspond almost exactly with those obtained from the Kelley table. (A mathematical method for obtaining the "proportion of differences" is explained at the end of the article.)

It should be emphasized that the reliability and intercorrelation coefficients used must be both realistic and appropriate to the group with which the tests are to be used. The spuriously high reliability coefficients resulting from the use of split-half or Kuder-Richardson formulae with speeded tests will give overestimates of the proportion of differences. Coefficients obtained from groups of greater or lesser homogeneity will yield distortions. Very often the coefficients given in test manuals do not satisfy these requirements, apparently having been selected more for promotional value than for authenticity.

Kelley, in his original article, applied the method to data from the first edition of the *Stanford Achievement Test*. Segel has used this method with the three parts of the *Gates Primary Reading Tests* and with the four parts of the *Gates Silent Reading Tests*. The average proportion of differences for these tests is not far from .25, which value is regarded by Segel as about the minimum degree of differentiation required for useful diagnostic tests.

The Kelley method was recently applied to two fairly well-

known batteries for which sufficient data are available. The *Detroit General Aptitudes Examination* produces three subscores entitled "Intelligence," "Mechanical," and "Clerical." The intercorrelations are all higher than .7 and the reliabilities,

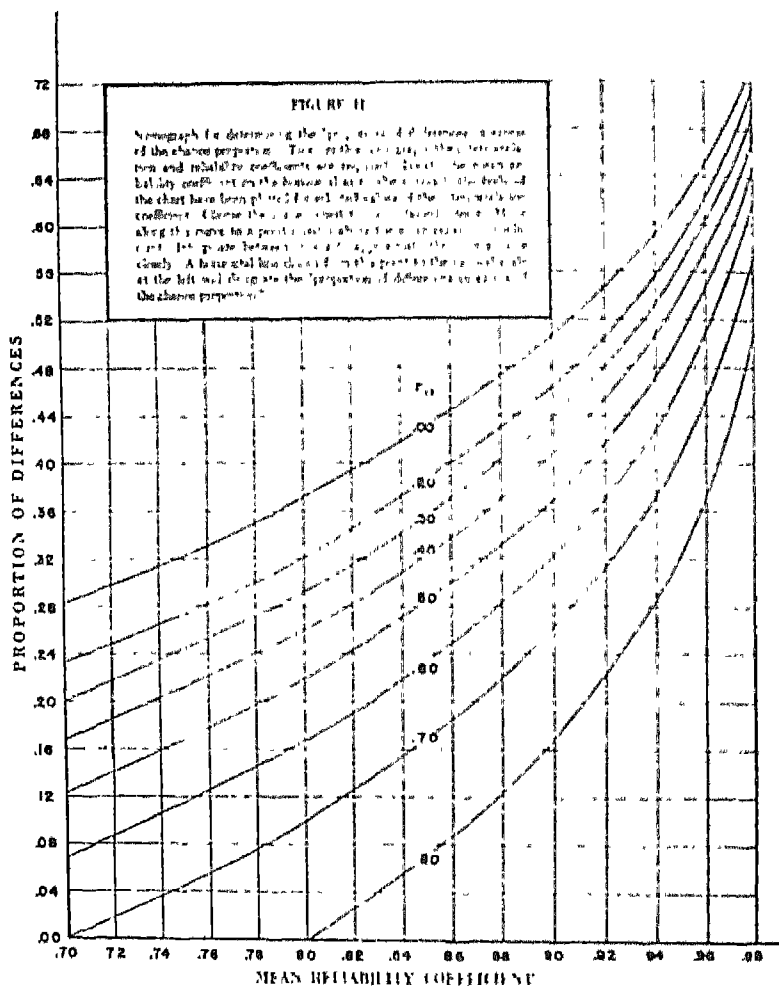


Fig. 2

above .8. The proportions of differences are .06, .07, and .21, none meeting Segel's requirement. The second test battery is the *Engineering and Physical Science Aptitude Test*, edited by B. V. Moore. There are six sections in this instrument with reliabilities ranging from .68 to .93. The 15 intercorrelations

range between .28 and .61. The Kelley index values are from .14 to .42 with six exceeding .25.

More extensive use of the method has been made by Bennett, Seashore, and Wesman in connection with the *Differential Aptitude Tests*.³ This series of eight tests was, from its inception, designed to produce a profile of scores useful in the guidance of high-school students. It was apparent that the battery must have good differential power to justify its use in the manner intended and the Kelley method proved of help throughout. The first step in this development was to study available test data to see which pairs of meaningful measures could effectively be used at the same time. Since the practical upper limit of reliability for short power tests is about .9, it was decided that the maximum permissible intercorrelation would be about .6 with lower values desirable. This approach clearly indicated that certain combinations of variables, such as arithmetic reasoning and arithmetic problems or synonyms and verbal analogies could not be expected to contribute importantly to the diagnosis of ability. On the other hand, it was somewhat surprising to find that spelling and grammar, which were originally considered as subsections of "language usage," were sufficiently independent to warrant individual consideration.

Having selected a promising set of tentative variables, the next step was to find item forms suitable to the ability levels intended, namely, grades 8 through 12, and producing the maximum reliability per unit of time. It is not feasible to discuss here the various item types tried, but the test authors were successful in producing a series of eight tests, none over thirty minutes in length with reliabilities ranging between .85 and .93 for a single grade of high-school boys and, with the exception of the *Mechanical Reasoning Test*, between .86 and .92 for a single grade of high-school girls. The intercorrelations fall between .06 and .62 for boys, and between .12 and .67 for girls. The respective mean intercorrelations are .38 and .39.⁴

³ *Differential Aptitude Tests* (Manual and eight tests). New York. The Psychological Corporation, 1947.

⁴ These data are taken from Section C of the Manual for the *Differential Aptitude Tests*

The application of the Kelley method to these data yields quite satisfactory results. The proportions of differences in excess of the chance proportion are as follows:

For boys, from .29 to .52 with a mean of .39.

For girls, from .20 to .48 with a mean of .34. Four values are lower than .25 and these all involve the mechanical reasoning test which has unsatisfactory reliability when administered to women.

In conclusion, the authors should like to say that their experience with the Kelley method leads them to recommend it highly both as a means of evaluating existing test combinations and as an aid in the construction of new test batteries

Mathematical Note on Computing the "Proportion of Differences"

Consider the two normal curves shown in Figure 1.

The equations of the two curves may be written as

$$y_1 = \frac{N}{\sigma_1 \sqrt{2\pi}} e^{-(x^2/2\sigma_1^2)} \quad \text{and} \quad y_2 = \frac{N}{\sigma_2 \sqrt{2\pi}} e^{-(x^2/2\sigma_2^2)}$$

Since the curves are of unit area $N = 1$

Let $\frac{\sigma_1}{\sigma_2} = k$ and let $\sigma_2 = 1$

$$\therefore \sigma_1 = k$$

The equations of the curves may then be rewritten as

$$y_1 = \frac{1}{k\sqrt{2\pi}} e^{-(x^2/(2k^2))} \quad \text{and} \quad y_2 = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}$$

To find the abscissa values of the points of intersection of the two curves:

$$\frac{1}{k\sqrt{2\pi}} e^{-(x^2/(2k^2))} = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \quad \text{or} \quad e^{-(x^2/(2k^2))} = k e^{-(x^2/2)}$$

Take logarithms of both sides (common logs have been used in this development):

$$-\frac{x^2}{2k^2} \log e = \log k - \frac{x^2}{2} \log e$$

$$\frac{x^2}{2} \log e \left(1 - \frac{1}{k^2} \right) = \log k$$

$$x^2 = \frac{2k^2 \log k}{\log e(k^2 - 1)}$$

$$\text{Since } \log_{10} e = .4343 \quad x^2 = \frac{2k^2 \log k}{.4343(k^2 - 1)}$$

x , the abscissa values of the points of intersection of the two curves, is given by:

$$x = \pm \sqrt{\frac{2k^2 \log k}{.4343(k^2 - 1)}}$$

k , the ratio of σ_1 and σ_2 , may be determined by applying the formulas shown under Figure I. The value of x may then be found by using the above formula. The areas under the normal curve for abscissa values of x and $\frac{x}{k}$ may be found from a table of area under the normal curve. The difference between the two areas multiplied by 2 is the "proportion of differences in excess of chance."

A STUDY OF CLIENT SELF-SELECTION OF TESTS IN VOCATIONAL COUNSELING

JULIUS SEEMAN
University of Chicago

It is a basic postulate of psychology that organisms react selectively to their environment. Stated more inclusively, the principle holds that within limits of structure and capacity, organisms tend to select from their milieu those elements which preserve their existence, foster their development, and maintain their integrity as separate living entities. Professional workers have been concerned with the application of this principle in diverse areas of human experience and have dealt with it at varying levels of discussion and demonstration.

Davis (3) has reported a long-term study designed to test the adequacy of self-selection of diets by infants. In this study, conducted with babies at the time of weaning, a representative assortment of widely used common foods was presented to the children in separate dishes placed on a tray. Each child was allowed to eat whatever foods he wanted in whatever order and quantity he desired. Davis concluded that the children not only ate adequate diets and grew in normal fashion, but that there was a marked absence of anorexia, behavior problems connected with eating, or need for cathartics at any time during the experiment.

The use of the self-selection principle has been reported in two descriptive studies, by Olson (6) with reference to nursery-school sleeping schedules and by Olson and Davis (7) in connection with development of reading ability and interest. In the nursery-school study, Olson described an experience at the University of Michigan Elementary School in which it was decided to follow the full implications of the self-selection principle and not require afternoon naps of all children, as had theretofore been done. The report stated that

The a priori arguments regarding the social effects of the example of non-nappers on those who still napped proved unfounded; some children napped regularly, others only when exceptionally tired. Teachers testify to the increased serenity of the children (6, p. 54).

The reading study was designed to observe the reading growth of children in a permissive environment in which each child could set his own limits with respect to the number and kind of books to be read. A wide variety of books was placed in the classroom; the teacher made no assignments to the children, acting only as the children needed and desired reading assistance. At the close of the term, evaluation disclosed wide variation both in the number of books read and the reading ability of the children. The authors sorted out the correlative factors of the differential growth rates—mental age, physical attributes of the children, and social and emotional maturity—and concluded that "When an environment is supplied which is stimulating, the child reacts to it in terms of his total potentialities for growth" (7, p. 78).

Kurt Goldstein, through extensive work with brain-injured men, formulated the principle thus:

Normal behavior corresponds to a continual change of tension, of such a kind that over and over again that state of tension is reached which enables and impels the organism to actualize itself in further activities, according to its nature

Thus, experiences with patients teach us that we have to assume only one drive, the drive of self-actualization . . . (5, italics in the original).

In the field of personal adjustment, Rogers has stated a philosophy of counseling which makes much use of the self-selection principle. In a statement of the underlying assumptions of this method he said:

We have known for centuries that catharsis and emotional release were helpful . . . But we have not known or recognized that in most if not all individuals there exist growth forces, tendencies toward self-actualization, which may act as the sole motivation for therapy. . . . The individual has the capacity and the strength to devise, quite unguided, the steps which will lead him to a more mature and more comfortable relation to his reality (9).

It is the purpose of the present study to illustrate the application of the principle of self-selection to another area of experience. The study takes as its base an exploratory report by Bordin and Bixler, describing a method of test selection in vocational counseling in which the client is given the responsibility for selecting the kinds of tests he wishes to take (2). In this procedure the counselor describes to the client in non-technical language the tests which the agency has available and the information which the client may secure from each test. The client himself chooses the types of information he wishes to secure. The counselor makes the technical choice of the test which will best yield the desired information but does not add other tests or otherwise modify the client's choices. The method which these authors describe is based upon two premises: first, that clients who come for vocational counseling may be in need not only of information which tests can supply, but also of personal reorientation with respect to their attitudes, feelings, and motivations; second, that

...clients can deal most effectively with their own feelings and attitudes when they are active participants in the interview process, when they are permitted to attack their problems on their own terms, and when they are permitted to choose their own directions in grappling with their own problems (2, p. 362).

Before proceeding with an examination of some results of this method of test selection, it may be well to carry a step further the theoretical basis on which its use is predicated. It is not a novel conception that clients who come to counselors because they have vocational problems have at the same time a whole complex of drives, attitudes, and motivations—that they are, in short, unitary reacting organisms whose vocational problems and personality organization are part of a single system of behavior. Although this principle has received a good deal of clinical acceptance, it has as yet been the subject of little experimental treatment.¹ Agencies have published classifications and frequencies of problems bearing the rubrics “educational problem,” “vocational problem,” “personal problem,” or combina-

¹ Recent contributions by Bordin (1) and by Pepinsky (8) have dealt with problems of psychological diagnosis in student counseling

tions such as "vocational-emotional problem." The latter classification, though descriptive and serviceable, takes into account only that truncated range of behavior where maladjustment is in evidence. But in accordance with the "unitary reacting organism" theory, it is postulated here that interaction between vocational problems and total personality adjustment exists in some form at all levels of personal organization. A working hypothesis which takes such interaction into account through the whole range of behavior is that *the manner in which an individual meets his vocational problem is a function of his total prevailing level of adjustment*. Thus, it is suggested that a stable, well-adjusted individual will by and large solve his vocational problem in a manner more direct and less neurotic than an individual who has typically met situations neurotically. Conversely, an individual who has long been insecure and fearful, who has denied his problems, or who has developed maladjustive mechanisms may be expected to reflect these tendencies in dealing with his vocational problem. Since a counselor's work brings him into contact with clients of many levels of adjustment, he must be aware of and make provision for the personal-adjustment concomitants of vocational indecision. The method of client self-selection of tests is employed on the basis that it provides the possibility of a growth experience for the individual with basic adjustment problems as well as for the stable, well-adjusted individual.

The use of this method of test selection raises a number of questions which may best be answered by research evidence. The present report will concern itself mainly with two of these questions. The first deals with the kinds of test choices the clients make. Counselors often spend considerable time training in the use of tests; is it to be expected that clients can make appropriate choices without prior training and within the span of one interview?

The second question, a more general and exploratory one, concerns the interview behavior displayed by clients during the test-selection process. How do the clients react to the test selection process and what kinds of use do they make of the interview during the course of this process?

The Experimental Client Sample

The data to be reported were derived from the initial counseling contacts of fifty clients who came to the University of Minnesota's Student Counseling Bureau for vocational counseling between December 1, 1946, and March 30, 1947. The clients were interviewed by either of two counselors² who used the test-selection method here described. Six of these interviews for each counselor were electrically recorded. The criterion of eligibility for admission of data into the experimental sample consisted of the requirement that the client be one who came voluntarily to the Bureau for vocational counseling. This eliminated clients who were referred primarily for adminis-

TABLE 1
Tests of Homogeneity of Experimental and Control Groups

Variable	χ^2 or F	df	Result
1. Number of reasons for coming to Bureau	χ^2 , 4.632	10	Not significant
2. Types of reasons for coming to Bureau	χ^2 , 9.610	15	Not significant
3. Veteran or non-veteran	χ^2 , 8.385	5	Not significant
4. Year in school	χ^2 , 6.096	10	Not significant
5. College	χ^2 , 3.221	5	Not significant
6. High school rank	F, 4.225	127; 5	Not significant
7. Amer. Council Test	F, 1.610	5; 159	Not significant
8. Age	F, 3.333	5; 164	Significant at 1%

trative purposes; it did not eliminate clients who may have had other types of adjustment problems, so long as their presenting problem included a vocational problem. The sample consisted of fifty consecutive clients who met the criterion for admission.

In order to determine whether this sample was representative of the general Bureau clientele, the experimental sample was compared with a control sample of 120 clients who were interviewed by four counselors.² In each instance, the hypothesis under test was that the clients of all six counselors were drawn from the same population with respect to the variable tested. The following table indicates the variables on which the samples

² The writer wishes to express appreciation to Mrs. Vita K. Birnberg, Mr. Charles W. Goulding, Miss Phyllis Morris, and Mr. H. Wilkes Wright of Minnesota's Student Counseling Bureau and to Mr. Ray H. Bixler of the Minnesota Psychiatric Institute for their cooperation in the study.

were compared and the results of the test of the null hypothesis. The X^2 test was used for enumerative data and the analysis of variance for continuous data. For the analyses of variance, the use of the Welch-Nayer test revealed that the assumption of homogeneity of variance required for the use of analysis of variance was satisfied in all cases.

Variables 1 and 2 represent results of a questionnaire filled out by clients just prior to the interview and designed to obtain an indication of the clients' stated purposes in coming to the Bureau. It will be noted from the table that in these respects as in all others but one the hypothesis is accepted that the groups were drawn from the same population. With respect to the significant variable, age, the mean age of the experimental group was 21.0 years and the mean age of the control group was 20.4 years.

Evidence Concerning Types of Test Selection Methods Used

Since the data in the present study are predicated upon the assumption that the counselors for the experimental group used a method of test selection which allowed the clients to make their own choices, it is relevant to examine and test this assumption. It should be noted in this connection that many subtle factors may be at work during the test-selection process. A counselor who is diagnosing the client's needs as the process goes on, or a counselor who feels that he has a stake in the client's selection of any particular test, might easily provide a number of clues, consciously or unconsciously, which might be perceived and acted upon by the client. Thus the question of whether or not the situation is really a free situation for the client comes back to the counselor's underlying intent and attitude toward the client and the situation. The best available judgment of this attitude and the test-selection method used by the counselors would be based upon direct observation of the counselor, so that voice inflections and other mannerisms might be taken into account. Since this procedure was not feasible, the verbatim records (typescripts) of the interviews were used as a basis for judging the test selection methods used.

These judgments were secured by submitting to qualified

judges³ six recorded interviews for each of the two experimental counselors and six recorded interviews for each of the four control counselors. The judges were asked to read each interview and classify the test-selection procedure used according to four categories, ranging from complete counselor responsibility to

TABLE 2
Judgments Concerning Test Selection Procedures Used by Each Counselor

Counselor	Test Selection Procedure				Total
	A	B	C	D	
A	12	2	0	0	14
B	0	0	0	14	14
C	0	4	8	2	14
D	2	3	6	3	14
E	0	0	0	14	14
F	12	2	0	0	14
Total	26	11	14	33	84

complete client responsibility for selecting tests. The categories were as follows:

- A. Counselor selected tests with no choices made by the client.
- B. Counselor took major share of responsibility for selecting tests, but gave client definite opportunity to add to or subtract from counselor's list.
- C. Client took definite share of responsibility for selecting tests; counselor took selective role by offering chiefly those tests which seemed to the counselor to be applicable, or took persuasive role in "selling" one or more tests to the client.
- D. Client was given the responsibility for choosing the tests. Counselor limited his function to describing the values and limitations of each test in a neutral, non-persuasive manner.

Four of each counselor's interviews were classified by two judges and two were classified by three judges, so that each counselor's interviews received fourteen judgments in all. The interviews were submitted to the judges without identifying the counselor who conducted the interview. Table 2 presents the results of these judgments. In this table Counselors B and E were the experimental counselors, and Counselors A, C, D, and F the control counselors.

The chi-square value for this analysis was 117.94 (d.f., 15),

³ The writer wishes to express appreciation to Mrs. Virginia H. Bixler, Dr. Edward S. Bordin, Dr. Herdis Deabler, and to Dr. Paul E. Meehl for their work in judging the typescripts.

their own tests did not on the average select a larger number of tests than those used by the control group, but that there was

TABLE 3
Proportion of Clients Choosing Tests Available for Actuarial Prediction

Test	Cases Appropriate	Cases Selected
Gen'l Scholastic Apt.	22	19
Coop. Math (for entering engineering students)	4	4
Johnson's Science	4	4
Coop. Algebra (for entering students in ag., forestry, home economics)	4	3
Moss Nursing	1	1
Gordon Fractions	1	1
Coop. Gen'l Sc. (entering nurses)	1	1
TOTAL	37	33

TABLE 4
Proportion of Clients Choosing Tests Available for Clinical Prediction

Test	Cases Appropriate	Cases Selected
Interest Test	49	48
Social Studies (for prospective majors in sociol., polit. sc., hist., econ.)	8	7
Natural Science (prospective majors in med., dent., biological sc.)	7	6
Clerical (business detail majors)	5	5
English, Expression (prospective jlsn., English majors)	4	3
Dexterity (pre-dent students)	4	4
Wesley Social Terms (entering Jr., business)	3	3
TOTAL	80	76

greater variability in the number of tests the experimental group selected.

Evidence of discrimination or non-discrimination in test choices may further be obtained by examining the nature of client choices for tests which are more useful in some scholastic sequences than in others. For example, a test in spatial relations would be more useful in assessing traits desirable for certain technical-scientific occupations than for most occupations in the social science areas. Accordingly, two sub-samples were drawn from the experimental sample on the basis of membership in two distinct groups of claimed occupational choices. Group 1 consisted of nineteen individuals in the following classifications: pre-college or beginning freshmen students in engineering, architecture, and agriculture or forestry, and students in the

TABLE 5
*Incidence of Selection of Spatial Relations Test in Two Client Sub
Samples Within Experimental Group*

	Group 1 (Technical Group)	Group 2 (Social Science Group)	Total
Selected test	14	4	18
Did not select test	5	12	17
TOTAL	19	16	35

pre-medical and pre-dental sequences. The second group consisted of sixteen Arts College students (freshmen and sophomores) whose designated majors were speech, journalism, teaching, economics, business, and personnel work. It is assumed that in general the first group would find a test of spatial relations more useful than would students of the second group, since the attribute of spatial visualization plays a more important part in course work, laboratory, and practice for the first group than the second. The hypothesis to be tested in connection with these groups is that there was no significant difference in the frequency with which the respective groups selected the test of spatial relations. The chi-square statistic was used to test this hypothesis. Table 5 contains the data which form the basis of this test.⁴

⁴ In the actual statistical computation, Yule's correction for continuity was made by subtracting the value .5 from the larger numbers and adding .5 to the smaller numbers, since the small frequencies in the cells would otherwise tend to make X^2 spuriously high.

The obtained X^2 value of 6.411 (1 d.f.) is significant at the 2 per cent level. At this level of probability the hypothesis is rejected and it may be said that there is evidence of differential selection between the two groups. Inspection of the table reveals that the technical group chose the spatial relations test with significantly greater frequency. This would appear to reinforce the notion that clients select the tests in a discriminating manner according to their needs.

TABLE 6
Test Selection Patterns of Ten Clients in Experimental Group

Client No.	College	Year	Fields Considered	N											
				Gen. Apt.	Interest	Personality	G E D	2, Soc	Clerical	Math	English	Science	Social Studies	Dexterity	Spatial Relations
				1	2	1	4	5	6	7	8	9	10	11	12
1	Pre-coll.		soc. sc.	x	x	x	x			x					
2	Engin.	Soph.	archit	x	x	x									
3	Adult Sp.		speech	x	x	x	x	x						x	x
4	Engin.	Jr.	archit	x	x	x			x	x	x	x		x	x
5	Grad.		history		x										
6	Pre-coll.		business	x	x	x			x	x					x
7	Engin.	Soph.	engin.	x	x	x									
8	Engin.	Jr.	writing		x	x					x	x			
9	Pre-coll.		home ec.	x	x	x			x		x				
10	Engin.	Soph.	advt		x		x								
TOTAL				7	10	8	3	3	3	3	2	1	2	3	1

For the reader who may wish to examine patterns of tests selected by the clients, the test-selection patterns of ten clients in the experimental group are presented in Table 6. The ten clients here listed represent every fifth client of the alphabetically arranged list of the fifty clients in the group.

Client Reactions During the Test-Selection Process

This section of the report will observe at closer range the behavior of the clients during the test-selection process. In the earlier report by Bordin and Bixler, the authors make the following observations regarding client behavior at the time tests are being selected: "As he [the client] tries to puzzle things out, perhaps struggling with anxieties about the possible adverse

results of taking a test, the counselor helps him to clarify his feelings and to overcome the obstacles to accepting himself" (2, p. 364). They state further that: "Each and every test is a possible stimulant to the client's discussion of that field or phase of his life" (2, p. 368).

A study of the recorded interviews indicates the manner in which this process of client self-exploration takes place during the test-selection process. For purposes of illustration, several excerpts are presented here. The excerpts were all chosen from the point in the interviews where the mathematics test was being described; this was done so that it might be possible to observe the varying ways in which different clients responded to the same stimulus. In the twelve interviews, five clients responded to the mathematics test with a simple positive or negative choice and seven clients used the test as a focus of discussion. Three of the excerpts follow.

Excerpt 1

- C: Now another kind of test we have here is a math test. It will give you an idea of your math background, and compare your own background with that of other students here.
- S: Well, I had high school algebra - and that's my math. So perhaps it would be advisable to take that and find out where I stand in math. Are they practical problems? Very simple algebra I can probably handle, but when it comes to figuring out problems and stuff like that--well, I'll take that.
- C: You want to try it, but you don't think you're going to be very strong in the reasoning.
- S: No, I don't, but then we'll find out.

Excerpt 2

- C: This math test would give you an indication of your background in math and also help predict how a person is likely to do in our College of Engineering.
- S: I'm a little scared of math. But (pause) that would give me an idea of where I stand now in my math background. Right?
- C: That's right.
- S: Well, I think I've got an idea of where I stand now in my math background; It's one of my weaker points. But I've always been told that if I work at it, I could do well in it. But I don't know.
- C: It's hard for you to know about what others say concerning your potentialities

- S: That's right. At times in math, I've done very well when I applied myself a little more. But then at other times I didn't do so well and it just seemed like I hated it. And as a result, well, I got mighty low marks.
- C: You really had some ups and downs in math.
- S: Yes, that's sure. Well, it seemed to make a lot of difference as to the instructor I had. Some would tell me that I—well, in junior high school I was told that I was quite hopeless with it, and that sort of discouraged me. So I let it go for a year and then went to summer school to—well, to do what I'm doing now, to find if I could do it or not. Well, I got a B in it. So my math background now, I don't think, I know just about where I stand in that, I think.
- C: You feel that the test wouldn't be necessary because you can size it up for yourself.
- S: I can size it up pretty well for myself, that I'm not very high in math.
- C: Umhum.
- S: Well, on the other hand, maybe it wouldn't be a bad idea to take it.
- C: You're a little undecided on that one, aren't you?
- S: Yes . . . so I'd know actually just how bad off I am.
- C: So even though you don't think you're going to do well on it, you'd like to take it.
- S: Even though I don't think I'll do well on it at all.

Excerpt 3

- C: Now we have two tests here that are connected with how a student is likely to do in agriculture or forestry. They're an algebra and science test.
- S: Oh, I'll take the science.
- C: O. K.
- S: That isn't physical chemistry, is it?
- C: It's pretty general. It doesn't sound so good now, does it?
- S: Well, not if it's chemistry and algebra because I don't seem to be at all interested; but I'll try it.
- C: Just thinking about those things makes you feel — .
- S: Yes, that's right. It just gives me an ill feeling thinking about having to take those subjects. I'm worrying about them constantly, and that doesn't do me any good.
- C: Umhum. And worrying about those keeps you from doing as well as you might in other things.
- S: That's right.
- C: (pause) Do you feel that you want to take this test or not?
- S: I'll take the science test. I don't want to take the algebra test, though.
- C: You feel pretty sure of that, hum?
- S: Yes, pretty sure.

These excerpts provide a study both in similarities and contrasts. All of the clients had feelings of inadequacy in mathematics, yet the degree of self-acceptance which they displayed regarding this inadequacy and the extent to which they struggled with the prospect of exposing themselves to a math test were unique with each client. Actually, the whole process of test selection appears to afford the client an opportunity to come to terms with his own feelings and attitudes and to examine these attitudes in an atmosphere where threat or evaluation by the counselor is at a minimum.

The final aspect of the present report deals with the reactions of the clients to the test-selection process itself. In this connection it may be said that the majority of clients who come to the Counseling Bureau for vocational counseling expect testing to be part of the process.⁵ Their prior knowledge of the nature of the tests varies, but the modal preliminary set of the clients is that they will be given a test or battery of tests uniform for all clients. When the counselor uses the test-selection method here under consideration, many clients give no indication of difficulty in working with this situation; others find it difficult to deal with a setting which differs from their expectations and find it hard also to take the responsibility that the self-selection process implies. The relative frequency of these alternative reactions may be estimated from the fact that in the twelve recorded interviews, eight clients undertook the self-selection method without evidence of difficulty or confusion and four clients revealed varying degrees of indecisiveness and dependence upon the counselor. Since these modes of reaction will be discussed at further length, it may be worthwhile to illustrate the operation of each as they occurred in the interviews. In Excerpt 4 the client appears to have no difficulty with the process, while in Excerpt 5 the client finds it difficult to select the tests.

Excerpt 4

C: This test is a test which gets at your ability to judge art principles. It's not a drawing test; it's a test where you

⁵ Of the 170 clients in both experimental and control groups, 145 checked as one of their reasons for coming to the Bureau the statement, "I want to take the aptitude tests so that I can find out what vocation to go into."

select in each case the better of a pair of pictures just to see whether you understand the art principles involved in the pictures.

S: I see. Well -that wouldn't have anything to do in the line of engineering drawing, though - it wouldn't help me out there, would it?

C: No, it won't predict how a guy will do in that. We do have a test which helps give you a prediction of that nature.

S: Well, I think if I were taking any drawing, it would be some kind of engineering drawing

C: You've pretty well ruled out other types

S: Yes.

C: This test (pointing to another one) does give some indication more along lines of blueprint, layout work, engineering drawing. It gets at your ability to visualize various sizes and shapes of figures, the kind of thing you need when you do a blueprint.

S: I want that one then.

Excerpt 5

C: This test will give you an indication of your general ability for college work. We find that it helps to predict how you're likely to do in our liberal arts college here. This second one is an intelligence test which will compare you with the general run of people; it'll come out with a comparison of your intelligence with the intelligence of the population in general.

S: Well, I'd just as soon take them. Is there any charge for them?

C: No, not for students here at college.

S: Well, I want to know. What do you think? I'd just as soon take any ones you suggest. Frankly--I mean, the more I can find out about myself the better I feel--without going too far, that is. I mean, not taking every single one on the list.

C: Yeah. That's basically why I'm going through these with you, so that you can get an indication of what you can learn about yourself; that may help you to be selective about the tests you want.

S: Well, anyway you think. I mean, I don't know too much about the tests. If you think I should have it, I'll take it. I don't know--I don't think--I don't care--if I have to take it I'll take it. If you think I should take it, I'll take it; put it that way.

C: You'd like to depend on my judgment in this.

S: Yeah, that's right. I mean, taking tests like this doesn't bother me.

C: Well, the information is going to come to you eventually and you'll be the one to use it. That's why I'd like to make this a cooperative venture and why I'm not just laying out a bunch of tests.

- S: I've always wanted to take that test. I think I should.
C: You'd like to find out where you stand—
S: Yeah, I think I would
C: All right, Would you like me to check it?
S: Yeah.

It will be observed that in Excerpt 4 the client apparently knew what he wanted and specified the type of judgment he desired. In Excerpt 5, the client found it difficult to accept responsibility for selecting the tests and attempted to place this responsibility upon the counselor. It is interesting to reflect upon the implications of this kind of reaction for the more inclusive adjustment patterns of the client. Two alternative possibilities present themselves. The first is that the client's indecisiveness is specific to the test-selection situation; the second is that such indecisiveness in selecting tests is associated with indecisiveness in other areas of behavior and perhaps simply a reflection of a more inclusive problem of adjustment. These alternatives may be formulated in terms of a testable hypothesis based on the use of the recorded interviews. Specifically, the hypothesis to be tested is "Do clients who show dependence and indecisiveness in selecting tests also reveal more indecisiveness in other phases of the interview?" In order to test this hypothesis, the twelve interviews were first placed in random order and a count was made of the number of client responses in each interview which indicated general indecisiveness or confusion of purpose.⁶ This procedure was carried out independently by the writer and another judge. It will be recalled that four clients showed indecisiveness in selecting tests and eight did not. The interviews were divided into two groups on this basis and the frequency of signs of indecision was computed for each group. Table 7 indicates the count arrived at by each judge and also the pooled count (average) on which the computation and test of significance were based. It appears

⁶ Indecisiveness regarding test selection was not included, since this phenomenon served as the criterion for distinguishing the groups. Examples of client responses revealing indecisiveness are:

A. S: Maybe I'm going at this thing the wrong way. I can't forever jump from one field to another because it'll never get me anywhere. I'll never get through if I keep jumping around.

B. S. Yes. I mean, I had always imagined that when you were getting ready for marriage that—well, it should be just the very ultimate. That is, if it does come to the day when she's walking up the aisle and I can look at her and feel exactly what I wanted and exactly what I expected, well, I think that would be a very good way to get started. You see what I mean—I should be a little more sold on the idea

that in these interviews the criterion "signs of indecision" was an objective one, since the rank correlation of the independent judgments was .98.

A preliminary F-test of homogeneity of variance indicated that the two groups were homogeneous in this respect; therefore it was possible to pool the sums of squares for the two groups in computing the t-value. For this analysis, the t-value of 3.282 is significant at the 1% level. The table indicates that the mean number of signs of indecision for Group 1 was 4.3 while the mean for Group 2 was 11.875, over two and one half times as large as that for Group 1. It may thus be

TABLE 7
Frequency of Incidences of Indecision in Two Client Samples

Interview No.	Group 1 <i>Clients Not Showing Indecision</i>			Group 2 <i>Clients Showing Indecision</i>		
	Count, Judge 1	Count, Judge 2	Average	Count, Judge 1	Count, Judge 2	Average
1	3	2	2.5	14	11	13.5
2	10	9	9.5	11	13	12.0
3	3	1	3.0	8	8	8.0
4	3	4	3.5	14	11	14.0
5	2	2	2.0			
6	12	12	12.0			
7	1	1	1.0			
8	1	1	1.0			
Total	35	34	34.5	47	48	47.5
Mean			4.313			11.875
Sigma			4.12			2.72

said that the clients who showed indecision in selecting tests also displayed a greater incidence of indecisiveness in other parts of the interviews not related to the test-selection process. To the writer it seems that these signs of indecision on the part of the clients who found it difficult to select tests were but a reflection of a more basic insecurity and confusion of purpose—an insecurity which made it difficult for them to work out solutions either for the test-selection situation or for the other problems they faced. This phenomenon of pervasive indecisiveness seems to be in accord with the concept of vocational counseling stated earlier—namely, that the manner in which an individual approaches vocational solutions is a function of his total prevailing level of adjustment. To the extent that this is so, the

phenomenon of ambivalence and conflict in making test choices offers the same challenge and potential for therapy as a problem in any other area of personal adjustment, and may well be considered an integral part of the process of counseling.

Returning now to the relation between the general principle of self-selection and the data educed in the present study, it may be said that the data here presented appear to be in harmony with the self-selection theory. This conclusion must be seen in tentative terms, since alternate theoretical explanations of the present phenomena were not ruled out. However, it does appear from the data that clients do take a free situation and utilize it in terms of their own needs.

Summary

This paper has presented a study of client self-selection of tests in vocational counseling. The method of test selection here considered is based upon the assumption that it provides the possibility of self-exploration and consequent growth experience for the client.

Two questions raised by the use of this method were here explored: first, what kinds of test choices do clients make? Second, what kinds of uses do they make of the interview during the test-selection process? Data obtained relative to these questions were as follows:

1. Clients selected tests available for prediction in 93.2 per cent of the possible cases.

2. The discriminating nature of the test selections was further indicated by the facts that the clients chose an average of 5.71 tests out of over twenty-five available tests and that in the case of the spatial relations test the group of clients in technical sequences chose the test significantly more frequently than did clients in social science sequences.

3. Client reactions during the test-selection process varied from client to client. The test presentation appeared to serve as a stimulus sufficiently unstructured in nature to evoke in the client those reactions which his own background and attitudes dictated.

4. Clients varied in their reactions to the test-selection process itself. Those clients who displayed indecision in select-

ing tests also revealed a greater incidence of indecision in other phases of the interview.

REFERENCES

1. Bordin, Edward S. "Diagnosis in Counseling and Psychotherapy." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 169-184.
2. Bordin, E. S. and Bixler, R. H. "Test Selection: A Process of Counseling." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 361-373.
3. Davis, C. M. "Self-Selection of Diets: An Experiment with Infants." *Trained Nurse and Hospital Review*, LXXXVI (1931), 629-634.
4. Douglass, H. R. *et al.* *University of Minnesota Studies in Predicting Scholastic Achievement, Part II*. Minneapolis: University of Minnesota Press, 1942.
5. Goldstein, Kurt. *The Organism*. New York: American Book Co., 1939.
6. Olson, W. C. "Self-Selection as a Principle of Curriculum and Method." *University of Michigan School of Education Bulletin*, XVI (1945), 52-55.
7. Olson, W. C. and Davis, S. I. "The Adaptation of Instruction in Reading to the Growth of Reading." *Educational Method*, XX, (1940) 71-79.
8. Pepinsky, H. B. *The Reliability of a Set of Problem Categories for Use in the Clinical Counseling of College Students*, PhD Dissertation, University of Minnesota, 1946.
9. Rogers, C. R. "Significant Aspects of Client-Centered Therapy." *American Psychologist* I (1946), 415-422.
10. Strong, E. K. *Vocational Interests of Men and Women*. California: Stanford University Press, 1943.
11. Williamson, E. G., Bordin, E. S., Freeman, E. M. and Johnson, P. O. *University of Minnesota Studies in Predicting Scholastic Achievement, Part I*. Minneapolis: University of Minnesota Press, 1942.

A NOTE ON RAVEN'S PROGRESSIVE MATRICES TEST¹

H. J. A. RIMOLDI
University of Chicago

Introduction

THE main object of this note is to give some information about Raven's *Progressive Matrices*, which may be useful for the practical psychologist. We shall first describe the test, refer to some of the statistical values obtained during the process of standardization (7, 8) and compare these values with those obtained by other authors. A description of its factorial composition as obtained by including the test in a large battery and a study of the results of the introspective data that we have available will help the reader to form an opinion about the psychological functions active during its performance.

The Test

Raven's *Progressive Matrices* is a series of sixty items divided into five sets: A, B, C, D and E, whose difficulty increases with a certain regularity from item 1 in Set A to item 12 in Set E. The test is non-verbal and suitable for measuring intelligence above six years of age (4).

Each page of the booklet contains in the upper part a pattern in which a part is missing. At the bottom of the page several numbered pieces of the same size and form as that of the missing part are arranged in two rows. Of all of these, only one completes the pattern.

The subject may record each answer by writing the number of the chosen piece on an additional piece of paper, or may either vocalize or point to the correct answer with the fingers or any instrument when the test is taken individually. There is an equivalent form of the test for blind persons in which the

¹ Copies of the test may be secured from H. K. Lewis and Co., Ltd., 136 Gower Street, W London W.C. 1, England.

patterns are replaced by carved pieces reproducing the same problems.

By this description of the test, it is obvious that it does not depend to any large extent upon the education or amount of information of the subjects, that it can be easily applied to foreigners, and to normal or physically or mentally defective children (6). The test has been used successfully by Tracht (11) in testing cerebral palsied with different degrees of physical impairment.

All of the drawings are clearly presented in black and white. The instructions are short and simple, and the subject is asked to find which of the pieces completes the larger pattern. He is asked to work all of the problems without hurrying. Furthermore, there is no time limit. The mean time for a population of 1680 subjects between 9 and 15 years of age was 38 minutes with a S. D. of 11.90 (8).

Statistical Values²

The maximum possible score is 60, there being one mark for each item correctly solved.

Table 1 gives the means and standard deviations for our subjects in different ages and different conditions of testing. We shall not give here all of the values obtained in our standardization of the test. Those interested in more statistical details are referred to the author's previous papers (7, 8).

By analyzing the percentile values given in the aforementioned articles and comparing them with those obtained by Raven (5) the reader will find out that in spite of different environmental and educational conditions there is a considerable agreement in the results. This agreement holds both for individual and group testing.

The differences at each age level between the lower and higher percentiles—either for individual or group testing—is quite constant in the Argentine results; and the curves for all of the percentiles obtained by means of group testing show a fairly even trend, apart from a few irregularities, probably related

² The populations mentioned herein were composed of children in Mendoza, Argentina, studied between the years 1942-1946. The form of the test here employed was the *Progressive Matrices 1938* (4).

to biological factors and to the effect of selection of the populations (7, 8).

The individuals at a higher level develop more regularly than those below the median in the ability or abilities required to solve the test. The less endowed individuals show a slower trend of improvement.

The values obtained from group testing are on the whole slightly higher than those from individual testing.

There is a drop in the mean number of problems solved from Set A through C, there is no significant difference between the means for Set C and D, and there is a final drop in Set E. This may be further explained later on, when studying item difficulty.

TABLE 1

Age	Individual testing		Group testing	
	Mean	S.D.	Mean	S.D.
7	17.64	6.54	—	—
8	20.34	8.31	—	—
9	25.40	10.58	19.32	9.18
10	23.59	9.96	24.92	11.60
11	27.83	8.83	28.82	10.49
12	30.37	10.67	33.45	9.98
13	33.58	7.46	35.90	9.59
14	32.19	7.03	35.61	9.65
15	—	—	38.59	9.57

Item Difficulty

A comparison of Halstead's results (1, Table 7) with those of Rimoldi *et al.* (7, Tables 1 and 8, Table 5) reveals that the items do not increase in difficulty according to the presentation of the test. It can be seen that while the items of the last set are the most difficult ones, those of sets C and D appear scattered through the whole scale of difficulty.

Comparing the over-all difficulty of the items with the order of the problems as given by the author of the test, it is evident that while sets A and E are better limited, sets C and D have a greater scatter.

Study of the Errors

By the study of the most frequent mistakes for each item (7, 8) and its comparison with the report of Halstead (1) on

the same subject, one cannot fail to observe a similarity between the findings.

On the whole the opinion of the testees is favorable toward the test. It is quite obvious that they are interested in the problems since only a few of them fail to complete the test. In the younger children when the problems become more difficult there is a tendency to give up, which sometimes manifests itself by the casual way in which they reach the solution or by their repetition of the same answer time and again, a tendency also reported by Raven (6) in mental defectives.

Perceptual ability is important in the first two sets of the test. Frequently the wrong answers seem to indicate a faulty perception of the figure-field configuration. In Set B the attitude of the subjects is more analytical, although strong gestalt forces influence the solution of the problems. Often the wrong answer in this as in the following sets, especially that of E, seems to indicate that the subjects tend to work the solution in certain fixed spatial directions.

Sets C and D require a more complex type of activity on the part of the subjects. For example, some report "seeing the figures animated of movement."

In Set E it is not infrequent for the subjects to reach the solution by employing clearly analytical and algebraical procedures.

In addition to the above-mentioned ways of solving the problems, it is found that some subjects resort to all manner of manipulation with their hands when working the difficult items, others are guided by esthetic reasons, and still others give as the correct answer the most complex of the possible replies—findings that have been previously reported by Halstead (1) and by Raven (3).

Experience indicates that, on some items like D8, B12 and E8, the subjects reveal a wide variety of reactions, amounting sometimes to an obvious blocking.

Factorial Study of the Test

The five different sets of the test were included as separate variables in a battery in which some tests of performance, memory, cancellation and so on were present (9).

The results of the factorial analysis should be interpreted in terms of the special test configuration. Probably if the test were included in other batteries, other factors would contribute as well to its variance.

There is one factor common to all of the sets of the Raven, and to them only. Although it is difficult to interpret it clearly, one may well regard it as similar to Thurstone's factor I (10), insofar as all of the items require the discovery of a rule or principle for their solution.

Sets A, B, and C are loaded in a factor described as the perception of relations in space necessary for the construction of a whole. The fact that Set A, and to a lesser extent Set B are here present, confirms our opinion of the influence of perceptual and gestalt forces in the solution of the first two sets. Sets C and D seem to require this activity to a lesser extent.

Set B has a small loading in factor C, which is related to the construction of configurations against disturbing forces.

Set E is present with a small loading in factor D, a factor which has been described as related to the perception of right and left. In the most difficult items the solution is sometimes obtained by pointing to the pieces according to their position more than to their value for the solution of the problems, a fact also described by Miller and Raven (2).

Set E is also loaded in a memory factor. Sets D and E have saturations in a factor related to speed of perception.

Summary

The comparison of the values obtained by the administration of Raven's *Progressive Matrices* in different populations, different countries and different testing situations, shows that there is a strong similarity between these different studies.

For the reasons adduced in the previous paragraphs we recommend its use when the psychologist is interested in employing a non verbal test, well graded in difficulty, with good discriminative value and applicable to a wide age range. Moreover, the test can be applied to normal and handicapped subjects.

The loadings of the different sets in the different factors related to perception, construction of wholes, memory, relations

of right and left, speed of perception and a non identified factor, seem to play a considerable part in the actual solution of the items.

As some of these factors have been identified as similar to those found in other studies (9) with different batteries and different subjects, we suspect that they are basic components of psychological dynamics, beyond the mere presentation of the problem.

REFERENCES

1. Halstead, H. "An Analysis of the Matrix (Progressive Matrices) Test Results on 700 Neurotic (Military) Subjects, and a Comparison with the Shipley Vocabulary Test." *Journal of Mental Science*, LXXXIX (1943), 262-215.
2. Miller, F. M. and Raven, J. C. "The Influence of Positional Factors on the Choice of Answers to Perceptual Intelligence Tests." *British Journal of Medical Psychology*, XVIII (1939-41), 35-39.
3. Raven, J. C. "The R.F.C.I. Series of Perceptual Tests: an Experimental Survey." *British Journal of Medical Psychology*, XVIII (1939-41), 16-34.
4. Raven, J. C. *Progressive Matrices*, 1938. Instructions for Use and Norms, 1939.
5. Raven, J. C. "Standardization of Progressive Matrices, 1938." *British Journal of Medical Psychology*, XIX (1941-43), 137-150.
6. Raven, J. C. and Waite, A. "Experiments on Physically and Mentally Defective Children with Perceptual Tests." *British Journal of Medical Psychology*, XVIII (1939-41), 40-43.
7. Rimoldi, H. J. A., Cortada, N. and Velasco, E. S. "Ensayo de Tipificación de Una Prueba Mental (Progressive Matrices de Raven)." *Publicaciones del Instituto de Psicología Experimental de la Universidad de Cuyo*, I (1945), 83-114.
8. Rimoldi, H. J. A., Velasco, E. S. and Buhner, L. "Tipificación de los Progressive Matrices de Raven." *Publicaciones del Instituto de Psicología Experimental de la Universidad de Cuyo*, II (1948), 1-25.
9. Rimoldi, H. J. A. "Study of Some Factors Related to Intelligence." *Psychometrika*, XIII (1948), 27-46.
10. Thurstone, L. L. "Primary Mental Abilities." *Psychometric Monographs No. 1*. Chicago: University of Chicago Press, 1938.
11. Tracht, V. S. "Testing the Cerebral Palsied with the Raven's Progressive Matrices." Unpublished.

INDIVIDUAL DIFFERENCES IN SPEED OF RESPONSE IN MENTAL TEST MATERIALS OF VARYING DEGREES OF DIFFICULTY¹

MERLE W. TATE

Hamilton College

BEGINNING with what became known as the "personal equation" controversy regarding error in recording stellar transits, individual differences in speed of simple and complex response have been a subject of perennial interest and inquiry. The amount of research and controversy indicates both the importance of the subject and its elusiveness.

The persistent investigation of the significance of speed differences has taken six main lines of inquiry, several of which connect at one or more points. The earliest studies were those of the psychophysicists and dealt with simple reaction time in connection with sensation and other simple psychological processes.

The use of reaction time as a measure of intensity of sensation led to a second type of inquiry which was concerned with the effect of various conditions or factors, both internal and external, upon speed of discrimination and decision. In these studies it generally has been observed that the speed of correct judgments tended to be greater than the speed of incorrect ones, although there has been some adverse evidence.

The development of correlation technique at about the beginning of the 20th century encouraged a third type of study, the chief concern of which was the relation of simple and serial reaction time to school marks, performance in various intellectual tasks, and imputed intelligence.

The introduction of standardized group testing following World War I and the almost universal adoption of time-limit instead of either work-limit or untimed procedure in group test

¹ This paper is a summary of the thesis presented by the writer in partial fulfillment of the requirements for the doctorate at Harvard University, June 1947.

administration raised an issue regarding the significance of speed of response in test performance and led to extensive investigation of whether speeded mental tests discriminated against individuals who "might be slow yet profound."

A little later Thorndike and his students advanced the hypothesis that the dimensions of intellect included speed, width, and altitude. The hypothesis motivated a series of studies of the relation between the three, studies which connect with several earlier investigations of the relation between rate of work and quality or accuracy.

The studies of the relation of rate to quality of work led naturally to the question of constancy of rate in complex response, i.e., the question of whether an individual who was fast in one intellectual task tended to be fast in others. As early as 1900 the existence of a special speed ability in certain associative and perceptive processes was postulated, but research on the point was desultory for the next twenty-five years. With the development of factorial analysis techniques, however, the possibility of speed as an independent ability became of immediate concern, and during the last two decades the majority of speed studies have directly or indirectly considered the possibility.

Despite the great number of investigations along each of the last four lines, few hypotheses regarding the role of speed in intellectual response have been generally substantiated. The evidence has been conflicting. The more discriminating the research, the more obvious it has been that trustworthy data are hard to obtain. It does appear, however, that three conclusions can safely be drawn.

When measures, experimental conditions, and sample size have been adequate no appreciable relationship between speed in simple and serial reaction time and test intelligence has been found. The studies reporting significant relationship have employed subjects extremely heterogeneous as to chronological age or have been characterized by inferior altitude tests, crude measurement, or small samples and inexact statistical test. It appears certain that reaction time and test intelligence vary independently.

Although there is some conflicting evidence, it may be con-

cluded that group tests are not necessarily damaged by the imposition of time limits. In general, the scores in limited and unlimited time correlate highly and the scores in limited time show satisfactory correlation with other criteria. Much of the research on the point has, however, been trivial. Time limits for most group tests are established experimentally. The correlations between limited and unlimited time scores, frequently cited in support of the conclusion that speed may be identified with altitude, warrant, in view of the overlap or the part-whole circumstance, only the prosaic conclusion that time limits do not vitiate particular group tests. The improvement of correlations between group tests, with time limits removed, and other criteria has in general been no greater than would be expected due to the lengthened test.

The great weight of evidence indicates individual consistency in speed of response to a wide variety of mental test material of a low order of difficulty. The hypothesis of such an independent ability in speed in dealing with easy tasks is well supported both by correlations between direct measures of rate of work and by analyses of the intercorrelations of scores on speeded tests. Little is known about the nature of the factor, although it commonly has been described either as an aptitude or as a set for perceptual speed.

The evidence regarding the relation of speed to accuracy or quality and to altitude is particularly confusing. Several investigators have been able to identify speed with altitude in particular functions, and several have found significant correlation between speed and general test intelligence. About the same number have found only chance correlation. The time spent on achievement or mental tests generally has been found to be inversely related to the cumulative accuracy scores; the correlations may tell nothing, however, except that it takes longer to make errors. Significant correlation between speed and accuracy in arithmetic computation, speed and comprehension in reading, etc., usually has been found, but here again speed and quality have been measured simultaneously without a control of accuracy effect. The significant correlation reported between speed measured in low difficulty material and altitude measured in supposedly similar material of higher diffi-

culty may be due to significant relation between the two. On the other hand if, as some investigators have concluded, the duller subjects orient themselves more slowly in speed tests the correlations may be meaningless.

The majority of the inconsistencies and contradictions in previous studies may be accounted for by different methods of studying mental speed, dissimilarity in units of measurement, unreliability of measures, inadequate methods of controlling motivation, difficulty, and accuracy effects.

If two or more individuals responded to the same test items with the same degrees of accuracy, their performances could be compared simply and precisely on the basis of speed of response. Unfortunately the experimental control of accuracy of response results inevitably in test items of low difficulty, indeed, if defined in terms of percentages failing, in zero difficulty. Mental tests consisting of such items cannot be thought of as measuring anything except speed, since there would be no way of determining what else they measured, and provide no basis for generalizations regarding speed in different functions or in tasks of greater than zero difficulty.

Little is known about the effect of inaccuracy on speed of response, but it is reasonable to suppose that inaccuracy, like indecision in less complex response, dampens speed. The supposition is well supported by the numerous reports of negative correlations between cumulative accuracy scores and total times spent on tests. The supposition clearly is subject, however, to a qualifying provision. If an item were unintelligible to an individual a speed score might indicate nothing more than the time at which the individual judged it respectable to guess at the answer or to omit the item.

Neither total time spent on a test nor cumulative accuracy scores on a speeded test of greater than zero difficulty permit control of accuracy effect and neither provide the data needed for examination of the reliability of the speed measure. Furthermore, both measures, since they are affected by time loss due to gross irrelevancies such as picking attractive items, are at best crude approximations of speed of response.

Such considerations suggest that in a study of mental speed the test items should be selected to tap several dissimilar abilities, cover a wide range of difficulty, elicit at least two correct

and two incorrect responses from each subject at each level of difficulty under investigation, and should be timed individually. Other desiderata include (a) moderate length items requiring on an average of, say, 10 to 50 seconds, since errors in timing and the time differences ascribable to different rates of reading the stimulus material and writing the responses would be proportionately greater in shorter items, while gross distractions and irrelevancies would more likely arise in longer, (b) "open" or completion items, since multiple-choice items likely would encourage guessing under time pressure, (c) items relatively independent of school experience to minimize familiarity and practice effects, and (d) subjects relatively homogeneous as to chronological age, school progress, race, and motivation.

The Method of Investigation

This study, designed in light of the above desiderata, was made to determine (1) whether speed of response to mental test items of varying degrees of difficulty, with accuracy effects controlled, differentiates between individuals and (2) whether there is a factor of speed independent of altitude and independent of the function in which it is measured. It will be seen that the questions are basic, for they relate both to the possible existence of mental speed as an independent ability and to the reliability of direct measures of it.

The study was limited to a sample of thirty-six high-school sophomores and juniors having a mean chronological age of 188.4 with standard deviation of 7.4 and a mean mental age, as measured by the *California Test of Mental Maturity, S-Form*, of 218.7 with standard deviation of 16.7. The sample appeared to be normal with respect to test intelligence and school achievement. It was selected at random from the writer's mathematics classes in the Belmont, Massachusetts, High School. All members of the sample were enrolled in the college preparatory curriculum. Fourteen were girls and twenty-two were boys. Sex differences in sentence completion accuracy, favoring the girls, was significant by the *t*-test at the 3 per cent level, but the other differences in speed and accuracy were such as would be expected in sampling from a population in which sex differences were in fact zero.

The desirability of a random sample of Belmont high-school

sophomores and/or juniors was judged less than the desirability of obtaining a sample of interested and similarly motivated individuals willing to give the three to five hours required by the tests and amenable to necessarily rather inflexible schedules. Although the sample appeared entirely representative of Belmont sophomore and juniors and representative of such students in any upper-middle-class, suburban high school, the population it specifies must be considered hypothetical.

Preliminary investigation had indicated that speed of response to typical arithmetic reasoning, "find the rule and complete the series" number series, CAVID type sentence com-

TABLE 1
Levels of Difficulty in the Four Tests

	Arithmetic Reasoning	Number Series Completion	Sentence Completion	Spatial Relations
Level 1				
Number of Items	7	12	9	11
Median Difficulty	6%	7%	11%	6%
Range	3%-11%	5%-11%	3%-11%	3%-11%
Level 2				
Number of Items	18	16	22	20
Median Difficulty	32%	33%	28%	33%
Range	17%-42%	17%-39%	17%-39%	17%-39%
Level 3				
Number of Items	17	16	17	16
Median Difficulty	56%	47%	50%	47%
Range	44%-61%	42%-61%	42%-61%	42%-61%

pletion, and Army Beta spatial relations items significantly discriminated between individuals. Although the abilities elicited by such items of a wide range of difficulty are by no means independent, they were judged to be sufficiently dissimilar to demonstrate the presence or absence of a general speed factor.

The majority of the sixty-four arithmetic reasoning and the sixty number series and sentence completion items finally used in the study were selected from various well-known mental tests²; in order to get the range and levels of difficulty desired, however, it was necessary to construct about twenty items for

² The Otis, Henmon-Nelson, ACE Psychological Examinations, Thorndike CAVD Scales, Kuhlman-Anderson, and the California Test of Mental Maturity, Long Forms.

each test. The spatial relations items were adapted from the Army Beta and the Kelley³ tests. These consisted of outline geometric figures to be divided into specified parts. All items had the benefit of try-out on a sample of nine individuals comparable to the subjects in the final sample. The levels of difficulty, in terms of percentages of final sample failing, demarked by the items are shown in Table 1.

The tests were given during the Christmas holidays of 1945 and the two week-ends immediately following. In connection with the perennial class-room question of whether time-limit tests are fair to all students, the study had been explained to the subjects. At the time of taking the tests, each subject was told that his score would be based upon both speed and accuracy and he was asked to do his best in both respects. Each test was prefaced with rather extensive practice materials.

The arithmetic reasoning and number series completion items had been typed on individual cards and were presented in sequence, the subject writing his answer and the time announced by the examiner on the answer sheets. It had been observed in the try-outs that the tendencies to re-examine items or to perseverate were obviated by having the subject record his own time while the examiner got ready to present and time the next item, the latter, of course, checking on the correctness of the entry.

The sentence completion items had been duplicated, and a screening device which could be slipped down the page was used in presenting them. The spatial relations also had been duplicated, and the stimulus material, i.e., the parts into which the figures were to be divided, was presented on cards. The time spent on each item was recorded as in the arithmetic and number series tests.

No attempt was made to influence methods of work in the arithmetic test. Some of the subjects performed more of the tasks mentally than others. Whether time differences due to figuring the answers to some of the harder items on paper substantially affected speed or accuracy was a question which could not be answered from the data. Some of the faster subjects

³ Truman L. Kelley. *Crossroads in the Mind of Man*. Stanford University, California: Stanford University Press, 1928. p. 190.

figured on paper, as did some of the less accurate. Several of the subjects worked some of the harder number series on paper, but in that test, too, there appeared to be no association between methods of work and performance. The tests were given in the order indicated in Table 1. The testing periods required by different subjects varied from about three to five hours. Although the majority of the subjects complained of being tired at the end of their periods, it is not likely, judging from research on the point, that speed or accuracy was materially affected by tiredness.

In scoring the tests, responses were counted right or wrong. The decision to score the sentence completion items in the "all-or-none" manner came after an investigation of the consequences of using a partial credit scoring system. The rank order correlation coefficient between the systems was .96, and, in the interest of simplicity and uniformity in adjusting speed scores for accuracy, the "all-or-none" scoring was used throughout. The reliabilities of the tests, scored for accuracy, estimated from odd-even halves were without step-up, arithmetic .82, number series .88, sentence completion .80, and spatial relations .86.

The subject's speed of response to an item was measured by the number of seconds taken to complete the item. The relation between speed as ordinarily defined and the speed scores employed thus was inverse, the scores being really direct measures of slowness. An omission before time was called was treated as an incorrect response. This procedure introduced an uncontrollable source of error, but the alternatives either of estimating speed scores for omitted items from subject and item mean times, or of deleting the items upon which there were omissions, were judged worse. The former would have resulted in a loss of independence, and the latter in the loss of most of the items of above about 40 per cent difficulty. The use of multiple-choice instead of "open" items probably would have eliminated most of the omissions, but for reasons already noted the latter were judged preferable. In a few cases, it was necessary to impose the three-minute time limit which had been set to speed the administration of the tests, but these arbitrary scores were not used in the analyses of speed differences.

The basic data for the study were the 244 speed and the 244 accuracy scores of each subject, 64 in arithmetic reasoning and 60 in each of the three other tests.

The Appropriate Unit of Measurement

The two basic assumptions supporting variance ratio statistical tests are normality and equal variability in the parent population(s). Inspection of the speed scores, expressed in clock or raw time in seconds, indicated that neither assumption likely was tenable. All of the 144 frequency distributions of the thirty-six subjects in the four tests were positively skewed. In the majority the skewness, characterized by β_1 values in excess of +2.5, was severe enough to raise doubt concerning the propriety of variance ratio tests.

The four distributions of each of two subjects among the six having the highest cumulative accuracy scores for the battery, those of two among the six having the lowest, and those of five among the remaining twenty-four were randomly selected for further investigation of the normality and homogeneity assumptions. The extent of correspondence between each of the thirty-six selected distributions and a theoretical normal distribution was evaluated by the familiar Chi-square criterion. The probabilities that the observed distributions would arise in sampling from normal populations were extremely small, twenty-five being less than .001 with $P < .00001$ on the aggregate in each function, and it was concluded that the samples of raw time scores were not of normal parentage.

The distributions of the common logarithms of the raw time scores were examined in similar manner. All yielded probabilities of .05 or greater, the probability on the aggregate of the eight arithmetic reasoning log time distributions being .46, number series .27, sentence completion .78, and spatial relations .19.

In addition to preserving independence of the speed scores and normalizing the distributions, the log transformations substantially contributed toward equalizing the within-subject variances. The L_1 criterion (4) indicated that the hypothesis of homogeneity of population variances, highly discredited by the raw time samples in all four functions, was much more acceptable ($P > .01$) for the log time samples.

The thirty-six raw time and thirty-six log time distributions which had been selected for examining the normality and homogeneity assumptions appeared to be representative of the remaining distributions, and it was concluded that log time was the appropriate unit of measurement. Accordingly, all speed scores were converted to common logarithms, rounded to two-place mantissas and, to eliminate decimals, multiplied by 100. The log time scores were used in all later analyses.

As a matter for speculation, the demonstration of the appropriateness of the log time unit for measuring speed of response may have implications beyond the scope of this study. Distributions of test performances, however measured, obviously are affected by the talent of the subjects and the nature of the items. But in view of the range of talent and the spread and types of items in the four functions under consideration—three of which are fairly dissimilar—the findings may have considerable generality. The evidence is intriguing and would seem to add support to the general hypothesis that raw solar time is not an appropriate unit for measuring time dimensions of living matter (3).

Significance of Individual Differences in Speed of Response With Difficulty and Accuracy Effects Controlled

The means of the log speed scores of correct and of incorrect response of each subject at difficulty levels 2 and 3 (see Table 1) provided unbiased estimates of his speed in responding correctly and incorrectly at those levels of difficulty. With the exception of four subjects⁴ who missed no spatial relations items at level 2, each subject thus had four scores in each test. Six of the subjects had only one incorrect response at level 2 in one or another of the tests; their estimated speeds of incorrect response at that level, based upon single scores, may have contained large error. Rough approximation of the standard errors attaching to the other scores indicated that they ranged from about 2 to 25 with the great majority lying between 5 and 10.

⁴ The four subjects were not included in the analysis of the variation of the spatial scores. The alternative would have been that of estimating the missing scores from individual, level, and accuracy means at loss of independence.

The scores were considered linear functions of the true speed of the subjects, speed ascribable to the difficulty of the items and to the accuracy of responses, and speed ascribable to the

TABLE 2
*Estimated Speed of Correct and Incorrect Response at Two Levels of
Difficulty in Arithmetic Reasoning*

Subject	Level 1		Level 2		Subject Mean
	Correct	Incorrect	Correct	Incorrect	
1	154.2	187.8	165.4	168.2	168.9
2	156.3	140.2	154.0	158.4	152.2
3	126.4	147.0	114.2	143.4	132.8
4	123.1	136.8	150.7	149.0	139.9
5	157.8	166.4	181.4	183.6	172.3
6	152.9	175.8	149.7	181.7	165.0
7	127.4	147.5	123.1	161.8	140.0
8	150.7	148.3	157.3	149.7	151.5
9	149.5	150.8	141.7	154.5	149.1
10	148.2	137.0	154.0	158.7	149.5
11	146.5	138.8	180.0	182.4	161.9
12	140.6	147.5	160.1	145.0	148.3
13	165.6	181.5	173.9	189.1	177.5
14	133.7	187.4	164.3	163.4	162.2
15	143.6	170.7	169.5	176.6	165.0
16	147.0	139.4	152.6	154.2	148.3
17	174.1	180.8	199.4	167.5	180.4
18	131.1	153.7	145.8	141.2	143.0
19	167.1	168.8	186.2	165.6	171.9
20	142.4	152.0	142.8	166.5	150.9
21	147.6	144.5	165.5	172.5	157.5
22	166.4	160.2	176.9	185.7	172.3
23	158.5	197.8	175.0	161.9	173.3
24	129.3	113.8	131.4	133.8	127.1
25	138.9	116.4	154.3	163.3	143.2
26	153.4	130.1	151.2	153.1	147.0
27	147.4	143.0	157.4	143.8	147.9
28	129.1	123.7	143.5	136.0	133.1
29	172.2	154.3	170.6	151.2	162.1
30	136.8	142.4	155.9	143.2	144.6
31	151.1	151.5	154.9	152.3	152.4
32	148.0	142.5	158.6	151.6	150.2
33	138.7	136.3	150.8	126.9	138.2
34	123.8	137.1	140.0	127.9	132.2
35	137.7	136.8	146.4	152.5	143.4
36	122.3	135.8	133.8	145.1	134.2
Means	Level 1, 148.1 Correct, both levels, 151.0 Incorrect, both levels, 154.0		Level 2, 159.6		Grand, 152.5

random effect or experimental error independent of subject, difficulty, and accuracy. They were classified in each of the four tests according to the scheme illustrated in Table 2. In

testing the significance of the individual or subject, difficulty, and accuracy factors the general statistical test of linear hypotheses developed by Kolodziejczyk (2) was used. That test, in situations like the present, is equivalent to the familiar Fisher variance ratio test of differences between means and, like the latter, assumes homogeneity of the residual or criterion variance in the population sampled. The assumption was tested by the L_1 criterion (5) and was found entirely tenable ($.80 > P > .20$) in all four functions.

TABLE 3
Analysis of Variation of Estimated Speed Scores

Function	Source of Variation	Sum of Squares	Degree of Freedom	Mean Square
Arithmetic Reasoning	Between Subjects	28020	35	800.6
	Between Levels	2752	1	2752
	Between Accuracies	317.6	1	317.6
	Residual	12990	106	122.5
Number Series	Between Subjects	27380	35	782.3
	Between Levels	13460	1	13460
	Between Accuracies	3898	1	3898
	Residual	9709	106	91.59
Sentence Completion	Between Subjects	18160	35	518.9
	Between Levels	12520	1	12520
	Between Accuracies	9077	1	9077
	Residual	13583	106	128.1
Spatial Relations	Between Subjects	10580	35	302.3
	Between Levels	5824	1	5824
	Between Accuracies	11290	1	11290
	Residual	12520	94	133.2

The results of the analysis of the variation of the speed scores are entered in Table 3. The variance ratios obtained by comparing estimates of population variance provided by subject, level, and accuracy mean squares to the residual or criterion mean square in each function are shown in parentheses in Table 4. Had it been possible to include all thirty-six subjects in the analysis of variation of the spatial scores, the ratios in that function almost certainly would have been augmented.

With the exception of accuracies in arithmetic reasoning, all differences were highly significant, the probabilities being extremely slight of observing such differences in sampling from a population in which the corresponding true differences were

zero. Close approximation of the probabilities were obtained by Kelley's (1) cube root transformation of the variance ratios. These are shown in Table 4.

The insignificance of the accuracy factor in arithmetic reasoning was surprising. Judging from the answer sheets, there had been a tendency to oversimplify some of the items and to work easier problems than those posed. Then, too, the arithmetic items were less easily checked than those in the other tests and may have elicited fewer attempts to replace doubtful answers with better ones. But whatever the explanation, the circumstance did not appear to introduce any sort of systematic effect on the performance of the subjects.

TABLE 4
Probabilities That the Observed Differences in Speed Would Arise in Sampling from a Population in Which the True Differences Were Zero

Function	Differences Between		
	Subjects	Levels	Accuracies
Arithmetic Reasoning	.00000000+ (6 536)	.00001 (22 47)	1 (2 593)
Number Series00000000+ (8.541)	.00000000+ (147.0)	.0000000 (42.56)
Sentence Completion	.00000002 (4 051)	.00000000+ (97.74)	.00000000+ (70.86)
Spatial Relations0001 (2 562)	.00000003 (43 72)	.00000000+ (84 76)

In the above analysis, the estimated speed scores (illustrated in Table 2) were means of specified subsets of observed scores and were therefore partially free from the effects of within-subject variability. As has been noted, these means were characterized by relatively large error. It was desirable to ask whether the variation of true scores or of subject means was significantly greater than the variation of observed scores about subject means. It will be seen that the question related to the reliability of the speed measures, since all questions of reliability may be reduced to the single one of whether estimated true-score variation between subjects is significantly greater than the variation of observed scores within subjects.

Due to varying numbers of correct and incorrect responses

at the two levels of difficulty, an analysis of observed scores classified as illustrated in Table 2 would have been a matter of great complexity. It was judged more practicable to adjust the observed scores for accuracy effect. Sixteen items were

TABLE 5
*Speed Scores Independent of Accuracy on Seven Arithmetic Items of
3 to 11 Per Cent Difficulty*

Subject	Item							Mean
	1	6	7	61	62	63	64	
1	115	93	93	88	99	126	133	106.7
2	98	147	132	132	128	116	121	124.9
3	85	84	62	73	64	77	72	73.9
4	110	114	95	123	115	113	106	110.9
5	114	141	126	108	98	105	144	119.4
6	82	122	111	117	95	102	139	109.7
7	84	81	79	112	110	117	131	102.0
8	146	100	101	85	97	107	124	108.6
9	117	100	118	84	84	86	81	95.7
10	104	100	101	96	104	105	107	102.4
11	91	81	88	100	97	121	94	96.0
12	95	94	122	100	89	95	95	98.6
13	101	118	88	102	103	98	112	103.1
14	110	97	118	129	135	100	85	110.6
15	95	95	84	74	95	58	94	85.0
16	100	86	102	99	93	96	99	96.4
17	133	104	96	104	130	129	97	113.3
18	64	102	89	121	80	81	75	87.4
19	109	105	99	99	84	131	110	105.3
20	115	84	100	109	79	96	76	94.1
21	98	85	92	73	121	86	103	94.0
22	155	113	106	97	118	104	126	117.0
23	121	132	125	120	95	93	99	112.1
24	75	63	76	80	111	111	81	85.3
25	75	67	78	93	94	91	81	82.7
26	84	103	109	88	86	80	80	90.0
27	95	107	107	94	92	121	93	101.3
28	82	105	100	88	88	77	81	88.7
29	104	119	142	130	113	124	97	118.4
30	115	112	102	129	110	77	97	106.0
31	117	112	92	100	102	102	94	102.7
32	83	90	115	86	107	108	112	100.1
33	88	107	106	111	92	90	111	100.7
34	101	91	96	89	97	83	89	92.3
35	91	98	93	96	90	107	108	97.6
36	58	61	66	68	93	83	68	71.0

selected at difficulty levels 2 and 3 in each of the four tests, and the effect of accuracy was cancelled, at a loss of only two degrees of freedom per item, by computing the residuals of the scores from the regression line of speed on accuracy in each item. To eliminate decimals and negative signs, each residual

was rounded to a whole number and added to 100. For purposes of comparison, since it is known that a speed factor is elicited by easy mental test materials, the speed scores on all of the items of level 1 difficulty in each test were similarly ad-

TABLE 6
Analysis of Variation of Speed Scores Independent of Accuracy at Three Levels of Difficulty in Four Mental Functions

Function	Level	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Variance Ratio
Arithmetic Reasoning	1	Between Subjects	37028	36	1028.6	4.769
		Within Subjects	43571	202	215.70	
	2	Between Subjects	106617	36	2961.6	7.415
		Within Subjects	202900	508	399.41	
	3	Between Subjects	157259	36	4368.3	8.225
		Within Subjects	269783	508	531.07	
Number Series	1	Between Subjects	66279	36	1841.1	5.358
		Within Subjects	127818	372	343.60	
	2	Between Subjects	116145	36	3226.2	6.458
		Within Subjects	253795	508	499.60	
	3	Between Subjects	105656	36	2934.9	6.757
		Within Subjects	220650	508	434.35	
Sentence Completion	1	Between Subjects	40077	36	1113.2	3.775
		Within Subjects	89048	270	329.81	
	2	Between Subjects	64914	36	1803.2	4.218
		Within Subjects	217195	508	427.55	
	3	Between Subjects	95190	36	2644.2	5.527
		Within Subjects	243049	508	478.44	
Spatial Relations	1	Between Subjects	89677	36	2491.0	4.198
		Within Subjects	200570	338	593.40	
	2	Between Subjects	79237	36	2201.0	3.621
		Within Subjects	308760	508	607.80	
	3	Between Subjects	47376	36	1316.0	2.456
		Within Subjects	272189	508	535.81	

justed. Due to the process, the speed scores became independent not only of accuracy but also of item, level, and function effect, the only variation remaining being then ascribable to differences between and within subjects. The classification of the adjusted scores is illustrated in Table 5.

The results of the analysis of the variation of the adjusted speed scores in each of the twelve sets are shown in Table 6.

All of the variance ratios shown in the right hand column indicated that the between-subject differences were highly significant. By Kelley's (1) cube root transformation the probabilities that the differences indicated by the variance ratios would arise due to sampling, if the true differences were zero, ranged downward from about 1 in 100 thousand to less than 1 in a billion.

It should be noted that the assumption of homogeneity of within-subject variance, basic to the tests of significance summarized in Table 6, was entirely untenable at level 3, sentence completion, and was in doubt (P about .05) in the majority of the other eleven cases. The effect of lack of homogeneity in parent populations is not completely understood; empirical studies indicate that moderate heterogeneity does not cause intolerable divergence from theoretical F distributions. In the former case, however, it appeared likely that the heterogeneity was severe enough to introduce substantial error.

The Independence of Mental Speed

The fact that the log speed measures at the three levels of difficulty differentiated sharply between the individuals in the population specified by the thirty-six subjects having been established, the questions naturally arose (1) whether speed of response at the different levels in the four functions was constant, i.e., whether an individual's speed at a particular level in a given function was indicative of his speed at other levels in the same or in different functions, and (2) whether there was significant relationship between speed and altitude.

The means of the observed log speed scores adjusted for accuracy provided unbiased estimates of speed of response of the thirty-six subjects at three levels of difficulty in the four functions. (The means entered in the right hand column of illustrative Table 5, for example, were the estimates based on the seven items at level 1 difficulty, arithmetic reasoning.) Being means of the residuals from within item regression of speed on accuracy, they were free from item, level, and function effect; and their variation in a particular function was due only to differences between subjects and to differences within subjects from level to level.

For each subject, then, there were estimated true speed scores at each of the three levels of difficulty in each of the four functions. The question of whether speed was constant over the three levels was investigated, assuming homogeneity of residual variance, by testing the null hypothesis that the differences between subjects were zero. In arithmetic reasoning the assumption of homogeneity of variance was in doubt ($P = .05$); in the other functions it was clearly acceptable.

The results of the analysis of the variation of the scores, with allowances made for the loss in degrees of freedom resulting from the correction of the observed log speed scores for accu-

TABLE 7
Analysis of Variation of Estimated True Speed Scores at Three Levels of Difficulty in the Four Functions

Function	Source of Variation	Sum of Squares	Degrees Freedom	Mean Square	Variance Ratio
Arithmetic Reasoning	Between Subjects	16862	35	481.8	6.820
	Within Subjects	4945	70	70.64	
Number Series	Between Subjects	16833	35	480.9	13.20
	Within Subjects	2551	70	36.44	
Sentence Completion	Between Subjects	11612	35	331.8	8.191
	Within Subjects	2836	70	40.51	
Spatial Relations	Between Subjects	11996	35	342.8	5.729
	Within Subjects	4189	70	59.84	

racy, are shown in Table 7. The probabilities attaching to the variance or F ratios in the right hand column were, by Kelley's transformation, all less than 1 in a billion, and the hypotheses that between subject differences were zero could be rejected with statistical certainty.

The question of whether speed was constant over the four functions at each of the three levels of difficulty was similarly investigated, the assumption of homogeneity of residual or within-subject variance being tenable ($P > .20$) at each level. The results of the analyses are shown in Table 8. The probabilities of obtaining the variance ratios in the right hand column, if population differences were in fact zero, were all less than .00004.

When the speed estimates at the three levels in the four func-

tions were considered jointly, providing twelve estimates for each subject, the constancy of speed throughout all levels and functions could be examined. In that situation, the ratio of between to within-subject variance was very great, with $F = 19.11$. The supporting assumption of equal within-subject variance, however, was questionable, with $P = .009$. Thus considered, there probably were significant population differences in variability as well as in mean speed.

The intraclass coefficients of correlation, or fractions of the total variance of the true speed estimates ascribable to between-subject differences at the three levels of difficulty were .66 in

TABLE 8
*Analysis of Variation of Estimated True Speed Scores in the Four Functions
at Three Levels of Difficulty*

Level of Difficulty	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Variance Ratio
1	Between Subjects	11446	35	327.0	2.861
	Within Subjects	12603	105	114.3	
2	Between Subjects	12481	35	356.6	3.476
	Within Subjects	10777	105	102.6	
3	Between Subjects	12156	35	347.3	2.767
	Within Subjects	13174	105	125.5	

arithmetic reasoning, .80 in number series, .70 in sentence completion, and .61 in spatial relations; while the coefficients for the estimates in the four functions were .32 at the first level of difficulty, .38 at the second, and .31 at the third. All of the coefficients were, of course, highly significant—their significance having been established by the variance ratio tests—but the former four were substantially greater than the latter three. It seemed probably, therefore, that there were both a general speed ability and a special speed ability linked to function operative in response.

The fact that speed tended to be constant at the three levels of difficulty in all four functions suggested that, since it is known that speed in easy materials is independent of altitude, speed in harder materials might be indifferently related to altitude. The intraclass coefficients of correlation between the scores at the three levels, averaging about .70, however, indicated that

about 30 per cent of the speed variance was not predicted by the individual speed factor. It was possible that there was substantial covariation of speed with altitude at the upper levels of difficulty.

In planning the study, it had been hoped to have twenty-four to thirty items of about 50 per cent difficulty in each test, half of which might be used in estimating speed and half, together with the less and more difficult items, in estimating altitude. It developed, however, that none of the tests contained more than seventeen of such optimum items; and it was necessary either to decrease the reliability of the speed estimates or to measure altitude with benefit of few 40 to 60 per cent items. The latter course was judged the better.

The frequency distribution of the cumulative accuracy scores on each of the four tests with the 40 to 60 per cent items deleted showed, as would be expected, marked bimodality. It was therefore necessary to employ a system of scaling the items on the basis of their difficulty so that an altitude score might be projected into the gap. Under the assumptions that the abilities measured by the tests were sufficiently normal in distribution to permit the use of sigma values as indexes of item difficulty, that any interval on the sigma scale might be divided equally between the items falling within the range of difficulty establishing the interval, and that the divisions of the scale were additive, a simple scaling scheme for scoring the items available for measuring altitude was devised. The scheme effected a marked reduction in the bimodality of the altitude scores. The reliabilities of the scaled altitude tests estimated from odd-even halves, without step-up, were .71 for arithmetic reasoning, .65 for number series, .75 for sentence completion, and .79 for spatial relations. The rank correlation coefficients between the altitude scores on the scaled tests and the cumulative accuracy scores on the whole tests were all over .90.

The altitude scores of the subjects were the sums of the scaled values of the items they answered correctly. These scores, each multiplied by 100 and rounded to units, are shown in Table 9. The speed scores entered in the table were the rounded estimates of true speed at the third levels of difficulty. Thus, so far as the measuring instrument was concerned, the speed and altitude measures were independent.

The correlation of speed with altitude in arithmetic reasoning was $-.070$, in number series $.008$, in sentence completion $-.025$, and in spatial relations $-.071$. The coefficients were all less than one-half of their standard errors, and speed and

TABLE 9
Altitude and Speed Scores of the Thirty-six Subjects in the Four Functions

Subject	Arithmetic Reasoning		Number Series		Sentence Completion		Spatial Relations	
	Altitude	Speed	Altitude	Speed	Altitude	Speed	Altitude	Speed
1	207	109	170	122	209	127	268	102
2	146	98	143	97	142	98	223	109
3	269	70	194	65	217	92	207	87
4	170	91	216	84	146	100	186	91
5	200	125	226	110	204	121	150	102
6	174	117	246	82	255	113	216	85
7	276	82	291	105	312	100	288	100
8	184	90	191	97	123	95	250	93
9	149	90	210	92	224	96	184	114
10	147	98	157	114	193	90	133	106
11	184	122	148	101	219	101	145	100
12	247	99	262	110	235	102	152	84
13	208	126	189	121	200	82	214	106
14	184	108	173	104	182	119	185	113
15	220	115	234	103	232	101	298	105
16	176	93	168	109	165	92	199	90
17	243	141	201	111	243	92	214	100
18	224	88	261	80	269	104	235	104
19	194	118	246	109	158	108	166	102
20	218	95	250	93	243	72	260	99
21	203	119	266	112	248	107	248	97
22	190	126	211	130	166	112	245	124
23	241	116	233	106	166	118	215	106
24	241	78	238	89	301	104	316	101
25	218	98	230	94	254	104	125	110
26	174	96	189	88	177	87	224	101
27	202	94	197	95	204	100	150	109
28	260	86	267	95	263	120	307	92
29	131	97	87	87	111	98	175	101
30	193	94	168	93	142	118	119	91
31	206	98	249	105	201	105	316	86
32	201	96	195	114	182	79	252	101
33	170	76	166	80	158	88	279	107
34	188	76	207	109	193	84	153	101
35	215	90	237	108	235	93	214	101
36	202	82	260	85	297	82	213	82

altitude in the four functions, as defined, were clearly independent in the population specified by the thirty-six subjects.

Summary

In this study two broad questions have been considered, (1) does speed of response to mental test items of varying degrees of difficulty, with accuracy effect controlled, differentiate be-

tween individuals, and (2) is there a factor of mental speed independent of altitude and independent of the function in which it is measured.

Data were obtained by measuring the speed of response of thirty-six high-school sophomores and juniors to each of sixty-four arithmetic reasoning, sixty number series, sixty sentence completion, and sixty spatial relations test items of about 3 to 90 per cent difficulty - difficulty being defined by percentages failing. The speed scores measured in seconds were converted to common logarithms, the conversion having the effect of decreasing the heterogeneity of variances and of increasing normality of the distributions sufficiently to permit use of precise tests of significance.

When level of difficulty and accuracy effects were controlled, the null hypothesis that the mean speeds of response in the thirty-six sampled populations were equal was discredited with confidence approaching statistical certainty. The mean speeds were characterized by relatively large errors, but the hypothesis was found no more acceptable when the maximum error variance was taken into account.

It was found, when accuracy effect was controlled, that the subjects who were fast in response to items at one level of difficulty tended to be fast at other levels, the individual effect upon the variation of the speed measures being highly significant ($P < .000000001$) in all four functions. Furthermore, the subjects who were fast in one function tended to be fast in the other three, the individual effect upon the variation of the measures in different functions, at each of three levels of difficulty, again being highly significant ($P < .00004$).

It also was found that the correlations between measures of speed, defined by mean speed of response independent of accuracy to sixteen items of about 40 to 60 per cent difficulty, with independent measures of altitude in the same function were no better than chance, being $-.070$ in arithmetic reasoning, $.008$ in number series, $-.025$ in sentence completion, and $-.071$ in spatial relations.

The intraclass correlations of the speed estimates, or the fractions of their total variance ascribable to individual effect, at three levels of difficulty were substantially higher than those

of speed estimates in the four functions. This fact meant that relative standings of the subjects in speed were less affected by changes in difficulty in a function than by changes of function. It was concluded, therefore, that in addition to a general speed ability independent of altitude, the mental test items probably were eliciting a special ability in speed linked to the function in which it was being measured.

Considerable evidence might have been marshalled in support of the hypothesis that there were significant differences in variability as well as in mean speed between individuals in their responses to sets of homogeneous items. It is certain that there would have been highly significant differences had raw time measures been used. The differences in variability of the log time measures were not, however, except in the harder sentence completion items of sufficient magnitude to invalidate the statistical tests used. That log speed of response, with accuracy effect controlled, discriminated sharply between individuals in the population specified by the sample and that there was an independent factor of speed operative throughout a wide range of difficulty appeared statistically certain.

REFERENCES

1. Kelley, T. L. *Fundamentals of Statistics*. Cambridge, Massachusetts: Harvard University Press, in press.
2. Kolodziejczyk, St. "On an Important Class of Statistical Hypotheses." *Biometrika*, XXVII (1935), 161-190.
3. LeComte Du Nouy, Pierre. *Biological Time*. London: Methuen Co., 1936.
4. Nayer, P. P. N., "An Investigation Into the Application of Neyman and Pearson's I_1 Test, With Tables of Percentage Limits." *Statistical Research Memoirs*, I (1936), 38-51.
5. Welch, B. L. "Note on an Extension of the I_1 Test." *Statistical Research Memoirs*, I (1936), 52-56.

A COMPARISON OF FOUR MASCULINITY-FEMININITY SCALES

JOSEPH C. HESTON
DePauw University

THIS brief paper is not an exhaustive study of the traits differentiating masculinity and femininity. Terman and Miles (10) have provided the definitive report in this area. Our present study is primarily an empirical evaluation of (a) the degree to which four different MF tests agree and (b) their respective capacities to distinguish, i.e., separate into the anticipated dichotomy, between men and women. A third outcome of this analysis, arising from the necessary discussion of the tests themselves, will be a useful summarization of some of the types of items demonstrating utility in characterizing sex differences.

The four masculinity-femininity tests examined are the MF scales derived from each of the following: (a) Strong's *Vocational Interest Blank for Men*, Form M (9); (b) Kuder's *Preference Record*, Form BM (6); Hathaway and McKinley's *Minnesota Multiphasic Personality Inventory*, Group Form (2 and 3); and (d) Heston's *DePauw Adjustment Inventory* (4). The selection of these particular tests for MF analysis was largely fortuitous, i.e., the data for these specific tests were readily accessible from the records of counselees enrolled in the third annual Educational Guidance Clinic under the writer's direction at DePauw University in June, 1947. This clinic, described elsewhere (5), provided the opportunity to collect the MF scores on these four scales under certain advantageous conditions. First, the tests were all taken within a span of two or three days, thus essentially eliminating the intervening lapse-of-time factor which often proves troublesome in comparison of such tests taken at different times and under differing conditions. Second, since the subjects were counselees who paid a fee to participate in an educational guidance opportunity, we can fairly assume the tests were approached under optimum

conditions of sincerity, honesty, and general motivation. This latter point is important because it has been demonstrated that MF scores are easily faked, if the subjects have been so requested (10, page 77).

The four MF tests thus available permit a particularly fortunate chance to evaluate MF scales based on *dissimilar* categories of items. Two of the tests, the Strong and the Kuder, are composed fundamentally of items measuring *interests*. The MMPI (used hereafter to denote the Minnesota Multiphasic) is composed of interest items *and* personality questions. The DePauw scale has items designed to elicit feelings and emotions of a more subjective nature, and thus shows similarity to the personality items of the MMPI. An obvious hypothesis, then, at the outset, is that the two *interest* tests should prove more closely related as one pair of MF scales. The two so-called *personality* tests should form another pair of agreeing MF scales, to the extent that personality questions enter into their item content. The determination of the extent of these correlations is thus one of our major objectives. Each of the MF scales is briefly described below to facilitate observation of similarities and differences in item categories.

Description of the MF Scales

The Strong Scale.—Form M of Strong's blank for men contains four hundred items; 304 of these items have been selected as discriminative between men and women, and are thus scored as the MF scale on this test. The scale used in this study was Strong's sixth revision, in which the items are given weights ranging from +4 to -4. The scores used here are the *standard scores* based on this sixth revision of the MF scale. The author provides norms (9) based on a total of 372 men and 372 women. Each sex group has approximately one-third high-school students, one-third college students, and one-third adults. The basic types of items have been summarized thus:

- (a) *Items preferred by males:* mechanical and scientific activities; physically strenuous and adventuresome activities; legal, political, and army occupations; selling activities; and certain forms of entertainment (smokers, billiards, poker, etc.).
- (b) *Items preferred by females:* musical and artistic activities;

literary activities; kinds of people (largely unfortunate and disagreeable), certain forms of entertainment (fortune tellers, picnics, full dress affairs, entertaining, etc.); clerical work; teaching; social work; certain school subjects (such as Bible, botany, etc.); and items relating to merchandise (such as looking at shop windows, buying, displaying, etc.) (9).

The Kuder Scale. The MF score for the *Kuder Preference Record* differs from the other three MF scales used in this study inasmuch as it is not based *directly* upon individual item differentiation. In determining the MF score on the Kuder test one first is obliged to score the test on the nine basic preference area scales. One then multiplies each of these nine area scores by specified weights, some positive and some negative, and determines the algebraic sum of the weighted scale scores. In this operation, masculinity is evidenced by high positive weights on three scales, namely Mechanical, Computational, and Persuasive and by high negative weights on three scales, Artistic, Clerical, and Social Service. Masculinity is also indicated by smaller negative weights on the three remaining scales—Scientific, Literary, and Musical. Kuder, in his *Manual* (6), gives the computational procedure to determine MF scores for adults and also another set of weights to determine MF scores for high-school students. Thus far, no such scores have been developed and published for college students. We therefore, in this study, used the computational procedures set up for *adults* and the scores used are the final values of the designated computational procedure divided by one hundred. For the adult group, Kuder provides norms for random groups of 101 men and 100 women. He also provides occupational group norms for certain specific male and female groups.

MMPI Scale—In the MMPI test, the MF scale is composed of 60 items from the total group of 550 questions. In neither manual (2 or 3) is there any summary of the type of items thus utilized to differentiate between the sexes. The authors state some of the items were inspired by Terman and Miles and that others are original and that the items were originally selected by comparison of the two sexes. As the test is usually scored, a high standard score on the scale indicates deviation of the basic interest pattern in the direction of the opposite sex. In the

present study the score on MMPI is derived by *subtracting* the obtained raw score from the maximum possible raw score of 60. This makes a high score here indicative of masculinity and a low score indicative of femininity, and keeps the anticipated correlations with our other three variables positive, because in the other tests high scores denote masculinity.

The writer has made an analysis of the items in the MMPI MF scale and finds 27 of the 60 items concern themselves with likes and dislikes of an occupational or hobby nature. Thus, men like science, hunting, forest ranger, soldier, contractor and dislike librarian, florist, nurse, "drop-the-handkerchief," love stories, poetry, dramatics, etc. The other 33 items comprise the "personality reaction" portion of the scale and can be grouped somewhat as follows:

- (a) Twelve items expressive of emotional feelings (where the masculine answer denotes *less* emotional feeling); thus the male indicates his feelings are not easily hurt, he doesn't fear snakes, he daydreams very little, family habits do not annoy him, and he feels he does not react more intensely than most people.
- (b) Another category of items concern reactions to people. This set includes seven questions which indicate that the male in general has less trust in people. A male is characterized by feeling that he should if possible pay back wrongs, that he must stand up for his rights, that most people are honest only because they fear to be dishonest, most people make friends because friends are useful, that it pays to know who to get next to, etc.
- (c) Another category of five items is composed basically of those indicative of sexual inversion, such as too much interest in own sex and too much preoccupation with sexual problems in general.
- (d) The nine remaining items are of a miscellaneous nature, difficult to group under any significant heading.

The DePauw Scale.—The *DePauw Adjustment Inventory*, which is thus far experimental and has not been released for general use, is composed of three hundred "personality" items and has been developed to score on six basic scales in addition to the MF scale. The six basic scales have been derived by item validation and no item scores on more than one scale. The MF scale, however, cuts across all the other six, and the items constituting it have been selected on the basis of their power to

differentiate between men's and women's answers. The population upon which the MF scale was originally constructed was composed upon two hundred men and two hundred women, all college freshmen at DePauw University.¹ Seventy MF items have been thus selected, composed of categories described as follows:

- (a) Sixteen items reveal the male to possess more emotional stability; i.e., he recovers more easily after being emotionally upset, his feelings are not as easily hurt, he is not easily startled, his mood is less influenced by others, and in general considers himself less emotional than others.
- (b) Eleven items indicate him to be more typically of the thinking introvert type; for example, he enjoys drawing conclusions from data, will question others' views and opinions, meditates about things in general considerably, thinks before action, does not take his work casually, enjoys "brain-teasers," and considers himself philosophically inclined.
- (c) Eleven items show the male has more confidence in himself; thus, it is not as hard for him to make decisions, he is quick and sure in his actions, he feels he can accomplish his goals, he plays his best against superior opponents, and does not wish his appearance were different than it is.
- (d) Eight items reveal that he is socially extroverted; he chats with clerks, he does not like evenings alone, he talks to strangers when traveling, he likes to speak in public, and has a ready answer when others make remarks to him.
- (e) Six items indicate the masculine trait of being more impartial and objective; he does not feel that other people are usually against him, people are not so bossy that he wants to be contrary, others do not offend him unwittingly, etc.
- (f) Seven items involve home adjustment; three show the male to be better adjusted and four show the female better adjusted.
- (g) The remaining eleven items are miscellaneous and difficult to place in any one descriptive category.

The Experimental Sample

The population available for this study is composed of all the enrollees in the 1947 Educational Guidance Clinic at DePauw University for whom scores on all four MF scales were available. Out of approximately ninety students enrolled in the clinic, we

¹ Miss Isabelle Loague carried out this MF item-validation research under the writer's direction.

had complete MF data available on thirty-four men and forty-five women. Most of these were 1947 high-school graduates coming to the clinic to secure educational guidance for their college freshman year. One or two of the men had military service experience intervening between high-school graduation and enrollment in the guidance clinic. The men range in age from sixteen to twenty-two, with a mean age of 18.3 years and an S.D. of 1.35 years. The women form a more homogeneous age group, ranging from sixteen to nineteen, with a mean of 17.9 years and an S.D. of .68 years. The group represents a fairly adequate sample of prospective college students interested in the various professions and in general liberal arts training.

The ability level of the entire group is considerably higher than average. On the *A.C.F. Psychological Examination* their median score is equivalent to the 71st percentile on the national norms. The *Cooperative English Test* reveals about the same degree of superiority, and the three *Cooperative General Achievement Tests* average slightly better than the 70th percentile for the group's mean score. We can generalize by describing the group as composed of students of better-than-average scholastic ability, but typical in most other respects of usual college freshmen.

It should be added that nowhere during the clinic program was any reference made to the fact that any of the tests which were given could be scored on an MF scale. Hence, we can assume the students were completely naive in this particular respect which, coupled with their generally excellent motivation in clinic participation, enhances the value of the data presented here.

The Experimental Data

In Table 1 we present the mean score and standard deviation for each of the MF scales, both for the experimental group from the clinic and the original groups providing the basis for the authors' norms. The experimental group of women corresponds fairly closely to the expected means and S.D.'s of the original norm groups. However, the experimental group of men, supposedly more or less a random sample, are consistently lower in MF score on all of the tests except the *DePauw Ad-*

justment Inventory than the original norm groups. Two factors might be suspected here. First, it is only the *DePauw Adjustment Inventory* norms that are based on a similar group of young college men. Strong's norms are based on high-school, college, and adult men combined; Kuder's norms are for adult employed men; and the MMPI norms are based on a group of engineers and unselected non-commissioned army men.

The other possible causal factor here is that we do not have a *random* sample of men, at least as far as occupational interests are concerned. A quick summary of the two or three high "group" areas exhibited by each of these men on the Strong test confirms this suspicion. Sixty-five per cent of these high

TABLE 1
Mean and S. D. of Each MF Scale for the Experimental Group Compared with Author's Norm

MF Scale	Group	N	Men Mean	S.D.	N	Women Mean	S.D.
Strong	Exp.	34	43.1	11.3	45	27.8	7.7
	Author	372	50.0	10.0	372	27.9	7.9
Kuder	Exp.	34	64.2	29.3	45	21.9	25.4
	Author	101	80.9	26.8	100	18.9	24.9
MMPI	Exp.	34	33.4	4.8	45	23.1	4.5
	Author	117	39.5	5.0	108	23.5	5.0
DePauw	Exp.	34	45.2	5.6	45	35.0	7.6
	Author	200	42.4	6.4	200	33.4	7.2

group scores are in Groups I, V, IX, and X of the Strong scale. It is precisely these groups, according to Strong's data (9) that exhibit *negative* correlations with the MF scale, ranging from approximately $-.12$ for Group I, to $-.60$ for Group X. It is easy to see, then, why this sample of men from the clinic, with interests concentrated in these four areas, would score lower than the expected norm on the Strong MF scale. The same explanation may hold for the Kuder test as well.

In Table 2 we have attempted to demonstrate the effectiveness with which each MF scale can dichotomize the experimental group into the two sexes. We recognize that the biological condition of being born male or female does not correlate perfectly with the possession of masculine or feminine interests. Nevertheless, it seems logical that an MF scale, built upon an item-analysis designed to select items differentiating between

males and females, should be fairly satisfactory in segregating the two sexes according to MF score. Accordingly, we have assumed, since in our present data masculinity is indicated by high scores, that the men should fall *above* the theoretical group mean (for the combined sexes) and that the woman should fall *below* that mean. Consequently, in Table 2 we have first computed the probable mean for the experimental group if the two sexes had been equally represented. Then, we have indicated the per cent of men on each scale that are above that theoretical mean, and the per cent of women that fall below. The average of these two can then be taken as an index as to how satisfactorily each MF scale separates the two groups according to sex. It is apparent that all of the scales do a fairly adequate job in this regard, three of them being about 78 per cent satisfactory

TABLE 2
Capacity of Each MF Scale to Dichotomize the Experimental Group According to Sex

MF scale	Theoretical Group Mean	% Men Above Mean	% Women Below Mean	Average % "Satisfactory"
Strong	35.45	76.5	80.0	78.3
Kuder	43.05	76.5	77.8	77.2
MMPI	28.25	85.3	81.4	84.9
DePauw	40.10	88.2	71.1	79.7

and the MMPI approximately 85 per cent satisfactory, making the latter the most effective in this respect.

The ability of each of the MF scales to distinguish between male and female subjects can be estimated by another approach, i.e., correlation. The standard biserial coefficient of correlation cannot be properly used in this connection because the customary biserial r assumes that even the dichotomous distribution does have a normal probability of distribution. Richardson and Stalnaker (7) have, however, suggested a formula for use in cases where there is no reason to logically assume normality of the dichotomized attribute. This formula produce what they call the point-biserial r . Applying their formula to the present data, we find the following point-biserial r 's between MF scores and sex:

Strong	.627
Kuder	.610
MMPI	.743
DePauw	.597

In each case we see a sizable positive correlation, thus demonstrating the ability of the scales in general to correlate with the respective sexes. Again, three of the scales (Strong, Kuder, and DePauw) exhibit approximately the same degree of effectiveness, with the MMPI demonstrating a somewhat higher degree of ability to separate the sexes.

A third method of evaluating the success with which each MF scale can differentiate between male and female is the use of what Peatman (8) has termed the *T Ratio*. This *T Ratio*, used here as synonymous with the older term, *Critical Ratio*, can be used as a test of significance for the difference between any two statistics. Here we have applied it to testing the significance

TABLE 3
 "T" Ratios of the Difference Between Sex Means for Data from Table 1.

Strong	6.80	33.45
Kuder	6.71	17.39
MMPI	9.75	23.91
DePauw	6.90	13.22

of the difference between the means earned by the two sexes on each of the MF scales. In Table 3 we show the *T Ratio* for the experimental group for the sex difference in means on each scale. In the parallel column are shown the corresponding *T Ratios* for each test as calculated by the writer from the authors' norm data shown in Table 1. The *T Ratio* for the experimental group is definitely significant for each of the MF scales, using the standard interpretation that a ratio of 3.00 is necessary to safely reject the hypothesis that the difference between the group means might be due to chance alone. It is obvious that factors other than chance have operated to produce the difference between the sex means in each case. It is again apparent that on the basis of the *T Ratio* exhibited, the MMPI is somewhat superior to the other three tests in ability to discriminate between men and women and the three others are again approximately equivalent to each other in this capacity.

The smaller ratios demonstrated for the experimental group than for the original authors' normative data may be explained

as due to two factors. First, we have already observed that our males in the experimental group were less masculine in their interest than one would anticipate on the basis of the authors' norms. Secondly, the experimental group is small, and as one increases N in a sample for computation of the T Ratio, this automatically increases the T Ratio, assuming the mean and sigmas remain constant. Thus if there had been one hundred men and one hundred women in the experimental group, exhibiting the same means and sigmas, the ratio would be raised to 15.6.

As indicated earlier in the description of these four MF scales, it is apparent that they approach the problem of measuring

TABLE 4
Intercorrelations Among the Four MF Scales
(based on 34 men and 45 women)

(MF Scale)	(a)	(b)	(c)	(d)
(a) Strong	—	.726	.686	.418
(b) Kuder	.726	—	.677	.414
(c) MMPI	.686	.677	—	.540
(d) DePauw	.418	.414	.540	—

masculinity-femininity from more than one angle. The Kuder and the Strong approach it definitely from the standpoint of interests and hobbies. The MMPI combines twenty-seven interest items with thirty-three personality items in its MF scale, and the DePauw scale is composed entirely of personality items. One wonders, then, how much agreement may be secured between MF measurements based on these different scales. In Table 4 we have answered this question by presenting the intercorrelations among all four of the MF variables.² The Strong and the Kuder show the highest correlation, as one would anticipate. The MMPI correlates almost as highly with the Strong or the Kuder. The DePauw is associated positively, but to a lesser extent, with both the Strong and the Kuder. The MMPI and the DePauw, having a common element of personality type items, correlate somewhat higher. The coefficients in Table 4 have not been corrected for attenuation

² Mr. Robert Brady contributed much of the correlational analysis in a research course for the author.

because the reliabilities of the Kuder and MMPI scales are not available.

As a further correlational consideration, it would be interesting to determine the effect if the MMPI scale were shortened so that the interest items were eliminated. We have therefore constructed a revised MF scale from the MMPI test composed exclusively of the thirty-three items that are *not* based on interests and hobbies. This abbreviated "personality MF scale" from the MMPI correlates with the DePauw .655, justifying the hypothesis that the personality items from the two scales are measuring more of the same thing than was apparent in the original correlation of .540. Further support for this approach to the problem is shown when one correlates this same abbreviated MMPI scale with the Strong MF scale. The initial correlation of .686 between MMPI and Strong now drops to .389, if one eliminates the interest items from the MMPI and uses only the personality items. This coefficient of .389 is very similar to the .418 exhibited between the DePauw and the Strong and the .414 between the DePauw and the Kuder. Thus a correlation of around .40 between MF scales based on these different types of items is apparently about what one should expect. Interesting support for this hypothesis is secured from Terman and Miles (10), for they report that the average correlation between any one of the seven parts of their test and the other six parts is .39. They also report their Part 4, which is made up of personality items (comprising anger, fear, disgust, pity, and wickedness) has an average correlation with the other six sub-tests separately of .39. Another scale which might be used to demonstrate the relationship between MF scales based on personality items is the "M" scale from Guilford and Martin's *GAMIN Inventory* (1). There are fifty-two items in their scale, with thirty-nine of them based upon emotional and personality reactions such as fears, annoyances, degree of confidence, etc. One can assume this MF scale would correlate rather highly with both the DePauw and the MMPI scale (minus the interest items).

Conclusions

Before we summarize this study into its major conclusions, one must bear in mind two limitations: (a) the relatively small

size of our experimental group, and (b) the apparently somewhat biased occupational interest distribution of our males, who consequently scored lower on MF than one would expect on all the tests except the DePauw. Now, under these limitations, one can conclude the following:

1. These four MF scales are about 80 per cent satisfactory in their capacity to place men above the mean score and women below the mean score. The MMPI is more effective than the other three in this respect.
2. The point-biserial r computed between each of these MF scales and the dichotomous category, men and women, in three of the cases is equivalent to approximately .60. The other test, MMPI, shows a higher point-biserial r of .74, thus is once again slightly superior.
3. Using T Ratio as a test of significance of the difference exhibited between mean scores earned by each sex, we find all four MF scales demonstrate the difference definitely due to something other than chance factors. The three tests Strong, Kuder, and DePauw show ratios of approximately 6.8 and again the MMPI is somewhat higher in this capacity with a ratio of 9.7.
4. In terms of intercorrelation, the two MF scales based on exclusively interest-type items show the highest association, $+.73$. The DePauw, based on personality items, shows the least correlation with the other three variables. The MMPI, being composed partially of interest and partially of personality items, correlates with the other scales to the extent to which such items are present in the other scales.
5. The MF measurements produced by scales based on dissimilar types of items may be expected to agree approximately in the neighborhood of $+.40$.

REFERENCES

1. Guilford, J. P. and Martin, H. G. *Inventory of Factors GMIN*. Beverly Hills, California: Sheridan Supply Co., 1943.
2. Hathaway, S. R. and McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory, (Revised Edition)*. New York: The Psychological Corporation, 1943.
3. Hathaway, S. R. and McKinley, J. C. *Supplementary Manual for the Minnesota Multiphasic Personality Inventory (for*

- Group Form*). New York: The Psychological Corporation, 1946.
4. Heston, Joseph C. *DePauw Adjustment Inventory (Experimental Edition)*. Greencastle, Indiana: Bureau of Testing and Research, DePauw University, 1945.
 5. Heston, Joseph C. "The Values of an Educational Guidance Clinic." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT* VII (1947), 299-307.
 6. Kuder, G. Frederic. *Revised Manual for the Kuder Preference Record*. Chicago: Science Research Associates, 1946.
 7. Richardson, M. W. and Stalnaker, J. M. "A Note on the Use of Biserial r in Test Research." *Journal of General Psychology*, VIII (1933), 463-465.
 8. Peatman, John G. *Descriptive and Sampling Statistics*. New York: Harper and Brothers, 1947.
 9. Strong, Edward K., Jr. *Vocational Interests of Men and Women*. Stanford University, California: Stanford University Press, 1943.
 10. Terman, L. M. and Miles, C. C. *Sex and Personality*. New York: McGraw-Hill Book Co., 1936.
-

AN EVALUATION OF FIVE FACTORS FOR PREDICTING THE SUCCESS OF STUDENTS ENTERING THE NEW YORK STATE COLLEGE OF AGRICULTURE

J. P. HERTTEL AND FRANCIS J. DIVESTA
Cornell University

As in many other colleges the counseling staff of the New York State College of Agriculture at Cornell University felt the need for measures which would supplement high-school averages in the prediction of college grades. On the basis of previous studies¹ of agricultural students, the *Ohio State Psychological Test*, the *Cooperative Natural Science Test* and the *Cooperative Mathematics Test* were selected and administered to all entering students, starting in 1942.² The results of a practical farm experience test which is required of all entering students were also available for study.

The group studied was composed of 324 male students enrolled as freshmen in the degree course between 1942 and 1946. These were students who had taken the three tests specified, as well as the New York State Regents examinations from which their high-school average was calculated. Students who substituted agricultural courses for certain required courses, such as English, in their freshman schedule during the war years were not included.

In order to study groups of students taking similar courses in their first semester, the total group was divided into the following curricula: preveterinary, agricultural science, and general agriculture. A comparison of the groups is shown in Table 1.

During the period studied 80 per cent of the freshmen

¹ "Studies in Predicting Scholastic Achievement," University of Minnesota, 1942; and unpublished studies in the College of Agriculture at the University of Minnesota, by L. H. Harden.

² The tests were administered and the computations in this study were made by the Cornell University Testing Service.

selected for the degree course ranked in the upper two-fifths of their high-school classes.

For purposes of analysis the first-term average was used as the criterion. Actual scores were used for each of the predictive measures. The regression equation was selected as the most suitable technique for determining the contribution of each factor toward the prediction of the college average.

The procedure³ entailed the computation of fifteen inter-correlations, twelve second-order partial correlations, six third-order partial correlations, one fourth-order partial correlation, five partial-standard deviations and five regression

TABLE 1
*Comparison of Freshmen Following Three Types of Curricula as Shown
by the Mean and Standard Deviation of All Factors Studied*

		Preveterinary		Type of Curriculum Followed				Total	
		Mean	Std Dev	Mean	Std. Dev.	Mean	Std Dev.	Mean	Std Dev.
College	Aver-	74.33	6.19	74.87	6.80	73.73	6.67	74.22	6.56
Coop.	Math.	82.11	11.43	83.68	11.60	82.08	12.14	82.37	11.80
Coop. Sci.	Test	57.99	13.68	61.68	14.01	58.36	12.68	59.19	13.43
Ohio St.	Psych	77.23	21.41	90.75	22.96	82.14	24.13	83.19	23.43
High	School	82.27	4.72	84.48	5.32	83.22	5.22	83.31	5.19
Farm	Av.								
Exper-	ience	22.51	14.45	12.71	13.03	22.75	14.13	19.86	14.62
Number	in	90		91		143		324	
group									

coefficients for each group and the total. The next step involved the utilization of the above data in the determination of the final regression equation and multiple correlation for each group and the total. It was assumed that linear correlation existed between all factors involved. This assumption was based on scatter diagrams made in previous studies. All calculations were made by machine. Original calculations were carried to four decimal places.

As shown in previous studies⁴ the results presented here indicate that high-school grades are more prognostic of college

³ Garrett, Henry E. *Statistics in Psychology and Education*. New York: Longmans, Green and Co., 1946.

⁴ Segel, David. *Prediction of Success in College* (Bulletin No. 15). Washington, D. C.: U. S. Office of Education, 1934.

scholarship than general mental tests or specific subject-matter tests. The highest association was found between the high-school average and college average, the correlation being .525 for the group as a whole (Table 2). When the groups were studied separately, a significant deviation was noted in the agricultural science group where the correlation is .398 as compared with .605 for the general agricultural group.

TABLE 2

The Inter-Relationships between First Term College Averages, Cooperative Mathematics Test Scores, Cooperative Natural Science Test Scores, Ohio State Psychological Test Scores, High-School Averages and Farm Experience of Freshmen as Shown by Correlation Coefficients

	Preveterinary	Type of Curriculum Followed		
		Agr. Science	Gen. Agr	Total
College Average and				
Coop. Math. Test	.188	.233	.294	.251
Coop. Nat. Sci. Test	.319	.259	.361	.321
Ohio St. Psych. Test	.398	.478	.460	.447
High-School Average	.538	.398	.605	.525
Farm Experience	-.103	-.058	-.029	-.049
Coop. Math. Test and				
Coop. Nat. Sci. Test	.139	.172	.212	.183
Ohio St. Psych. Test	.164	.364	.294	.281
High-School Average	.273	.275	.349	.310
Farm Experience	.104	.055	-.171	-.045
Coop. Nat. Sci. Test and				
Ohio St. Psych. Test	.494	.444	.456	.471
High-School Average	.297	.293	.474	.381
Farm Experience	.037	-.259	.188	-.017
Ohio St. Psych. Test and				
High-School Average	.344	.559	.536	.513
Farm Experience	-.109	-.021	-.182	-.172
High-School Average and				
Farm Experience	-.205	-.066	.038	-.095

Included in the high-school average there are relatively few science grades. The freshmen following the various curricula in agricultural science took a much higher proportion of college science courses than the other groups studied. Hence the high-school average was not as good a predictive factor for the agricultural science group. This is brought out more forcibly in Table 3 when the influence of other factors studied was held constant.

The inter-correlations show that the high-school average was associated most highly with the *Ohio State Psychological*

Test ($r = .513$) and least with the *Cooperative Natural Science Test* ($r = .381$) and the *Cooperative Mathematics Test* ($r = .310$). There was practically no association with farm experience except in the preveterinary group where there was a slight negative correlation of $-.205$. This may be explained by the fact that among students with little farm experience, only those with very good high-school averages are selected for the preveterinary course.

The second best predictive factor was the *Ohio State Psychological Test* which correlated $.447$ with the criterion for the whole group. The intercorrelation with the high-school

TABLE 3
*The Relationship Between College Average and Each Predictive Factor Studied
When The Influence of The Remaining Factors is Held Constant
as Shown by Partial Correlation Coefficients*

Holding remaining factors constant	Preveteri- nary	Type of Curriculum Followed		Total
		Aggr Science	Gen Agr	
College Average to Coop. Math Test	.547	.561	.090	.075
College Average to Coop. Nat. Sci Test	.098	.030	-.010	.002
College Average to Ohio St. Psych. Test	.209	.287	.197	.208
College Average to High School Average	.440	.172	.445	.353
College Average to Farm Exper ience	-.044	.058	-.047	.017

average was $.344$ in the preveterinary group compared to $.559$ in agricultural science and $.536$ in the general agricultural group. This may be a result of the homogeneity of the preveterinary students due to more rigorous selection.

Of the two tests designed to measure specific subject matter the *Cooperative Natural Science Test* correlated more highly with the college average ($r = .321$) than did the *Cooperative Mathematics Test* ($r = .251$). The correlation between these two tests was relatively low ($r = .183$).

When each predictive factor was studied with the influence of all other factors held constant, only the high-school average and the *Ohio State Psychological Test* proved to be of any real importance (Table 3). The fourth order partial correlation of the college average to high-school average was $.353$ and to

the *Ohio State Psychological Test* .208 for all groups combined. The *Cooperative Mathematics Test* and the *Cooperative Natural Science Test* were so highly associated with the *Ohio Psychological Test* and with the high-school average that their separate influence was negligible in predicting the first-term college average of these students.

The authors were surprised to find the relatively low association between the college average and the *Cooperative Natural Science Test* as evidenced by the zero order correlation (.259) and the fourth order partial correlation (.030) for the agricultural science group. The test did not appear to predict

TABLE 4
Relation of Actual First Term College Average of Freshmen Following One of the Curricula in General Agriculture to Average as Predicted by High-School Average, Ohio State Psychological Test and Cooperative Mathematics Test Scores

Predicted Average*		Actual Average				
First term	Number	90-99 Number	80-89 Number	70-79 Number	60-69 Number	60 & Below Number
90-99	0	—	—	—	—	—
80-89	24	1	13	10	0	0
70-79	100	0	12	59	27	2
60-69	19	0	0	8	10	1
Below 60	0	—	—	—	—	—
Total	143	1	25	77	37	3

* Formula— Predicted average = (Coop. Math. test score \times .04) + (Coop. Sci. test score \times .00) + (Ohio State Test score \times .05) + (High School average \times .63) + 14.91

scholastic achievement in college science courses for these students.

For all practical purposes it appears from Table 3 that the preveterinary group might have been combined with the general agricultural group. Those students following the more specialized curricula classified as agricultural science demand specialized attention when an attempt is made to predict their college average.

On the basis of the data mentioned, regression equations were developed for each of the specialized curricula. The multiple correlations were .55, .51, and .64 for the preveterinary, agricultural science, and general agriculture groups respectively. A multiple correlation of .57 was obtained for all

groups combined. A comparison of the actual and predicted scores of the general agricultural group is shown in Table 4.

Fifty-three per cent of the students who were predicted to make averages between 60-69, 59 per cent of those predicted between 70-79 and 54 per cent of those predicted between 80-89 actually fell in these groups.

Confronted with the problem of admitting students on the basis of the predictive factors studied here, an admissions officer in a college of agriculture would find it difficult to admit only those students who would make a 70 average or higher.

Although the multiple correlation of .64 appears statistically and graphically satisfactory, it still has serious limitations when put to practical use.

The difficulty of predicting college achievement increases with the amount of selection which has taken place in the admission of the students studied. In a group of college students selected because of their good high-school grades most of the causes of failure are probably factors which are not measured by existing tests.

Superficially, poor health, lack of self-discipline, inability to organize time and materials, too many extra-curricular activities and too much outside work appear to cause failures. On the other hand, most counselors have observed students who have succeeded in spite of similar handicaps. This difference might be accounted for if we were able to measure the factor commonly called drive or motivation.

A science test which would have a high correlation with college science grades and a relatively low correlation with the *Ohio State Psychological Test* would also contribute to an improvement in present predictive possibilities for students in the College of Agriculture.

Summary

These data support previous findings that the high-school average is the most important single factor in the prediction of the college average. The most important test of the battery used for prediction was found to be the *Ohio State Psychological Test*. The *Cooperative Mathematics Test* and the *Cooperative Natural Science Test* were negligible factors since they were

statistically dependent upon factors already accounted for by the *Ohio State Psychological Test* and the high-school average. Although preliminary studies showed that farm-reared boys were more successful than those who entered with no farm experience, the farm experience rating used in this study showed no significant association with the first-term average.

The highest multiple correlation of .64, although significant statistically, was found to be of limited value in predicting the success or failure of individual students.

Future studies attempting to predict college averages should include a better measure of aptitude for college science, a measure of motivation and in general, measures which correlate high with the criterion and low with each other.

•

A MEANINGFUL RECORD OF TESTS

EDWARD C. ROEBER

University of Missouri

THE recording of test scores on cumulative records and report forms is a tiresome occupation at best; nevertheless, the materials so recorded become a permanent record. Frightful predictions or evaluations are often based upon the symbols recorded on such records, the tests themselves having long since been destroyed. The exact situation becomes more apparent to anyone privileged to examine schools' cumulative record files. The writer has within the past few years discovered records which contained only the name of the test, the raw score, and the percentile rank. In some cases the name of the test was given as a meaningless abbreviation, viz., "Otis." It is quite evident that Otis has published several tests and several forms of each test. Counseling upon the basis of such information would be extremely hazardous for the trained counselor, who would probably reject such data as incomplete, without placing it at the disposal of many teachers who are inadequately trained for interpreting standardized tests.

The cumulative record which adequately records essential test data must still be in the process of formulation since it has not appeared on the market. What might such a record contain if it were to meet the standards of "safe-and-sane" counseling? It would seem that the following bits of information were a prerequisite to the understanding and interpretation of test results from test records:

1. *Name of Test:* The name should be as complete as necessary to distinguish it from *all* other tests. If it has been revised, this fact should be noted by the date of revision.
2. *Level and Form of Test:* Tests which have various levels, usually expressed as alpha, intermediate, elementary, primary, secondary, adult, etc., should have this fact recorded. The same holds true for the form of the test.

since, in spite of the comments of some test authors, there appear to be differences between the forms of the same tests.

3. *Raw Score*: It is rather standard practice to record this information so no comment is necessary.
4. *Percentile Rank, Standard Score, T-Score, Mental Age*, etc.: Again this information is adequately recorded on most records.
5. *Norm-group or Basis of Comparison*: A raw score is meaningless without comparing it with the performance of some group. In the case of a few tests, the norm groups are not as appropriate as the situation demands. It is, therefore, essential that the norm group be recorded since a test score will assume its true meaning when its norm group has been identified. Furthermore, additional significance might be added to scores by recording the mean and standard deviation of performance for the norm group.
6. *Probable Error of Test or Parts of a Test*: For years test authors have included probable errors of their tests in their test manuals. It would be interesting to discover how many teachers who interpret intelligence tests, for example, know the meaning of a probable error. Bingham¹ says, "When interpreting a person's performance, obtained scores should not be thought of as a point on the scale, but rather as a band or zone about that point; . . ." If this is the case, the present methods of recording a score as a point rather than as a band has only encouraged teachers to accept test scores as absolute. It is time that this practice be changed for the good of pupils. If necessary, the band should be recorded instead of a single score.
7. *Test Administrator*: Chance errors in the administration of a test may widen the band of values which are represented by a score. These errors may reside in the test administrator himself—thus the emphasis upon the inclusion of this information. Perhaps of equal importance would be some record of who scored the test and whether the test was re-scored to catch any errors.

¹ Bingham, W. V. *Aptitudes and Aptitude Testing*. New York: Harper and Brothers, 1937, pp. 253-254.

8. *Date of Test Administration*. This item may be needed in checking the age or grade when the test was administered. It is, therefore, more of a convenience for the counselor.
9. *Unusual Test-Administration Conditions*. This type of information assumes high importance when checks are made as to why supposedly similar tests reveal such startling discrepancies. There is ample evidence to show that physical condition, home conditions, and other emotional factors, have their effect upon any individual's performance. Furthermore, *rapport* between the test administrator and those taking a test is essential for the best performances. Besides the individual himself there are physical conditions, such as heat, humidity, extraneous sounds, and other conditions in the testing room which will have their effect upon test performance. A test interpretation which ignores the individual's condition and testing conditions is not valid—thus there is a need for recording this data.

The relatively complete information, suggested above, may not meet with the approval of those who record test data. Teachers, though, with such information are apt to avoid absolute statements regarding test results. And, if such an attitude were put into practice, it might avoid a possible reaction against all tests because of amateurish interpretations. Tests, assuming many imperfections, are better than chance. They are tools in much the same manner as chisels are the tools of carpenters. A chisel in the hands of an amateur is not the same tool in the hands of a master carpenter—and so it is with tests. Tests in the hands of an amateur can do much harm. In order to avoid an adverse attitude toward all tests and to avoid a recession to pre-war attitudes towards the testing movement, every precaution must be taken to avoid unscientific test interpretations. The proper reporting of test results is one step in that direction.

Summary

1. Present school records of test data are generally inadequate.
2. The following items are essential to adequate test report-

ing on cumulative or other records: (1) full name of test; (2) level and form of test; (3) raw score; (4) percentile rank, standard score, T-score, etc.; (5) norm group or basis of comparison; (6) probable error of test; (7) name of test administrator; (8) date of test administration; and (9) unusual test-administration conditions.

3. The testing movement may suffer another recession unless those who interpret tests are equipped with complete information on test records.

THE CONTRIBUTORS

George K. Bennett—Ph.D., Yale University, 1935. Psychologist, Blythewood Clinic, Greenwich, Connecticut, and Connecticut State Vocational Rehabilitation Service. Director of Test Division, The Psychological Corporation, 1936—. During the war: Member, War Relocation Authority; Panel of the Office of Scientific Research and Development. President, The Psychological Corporation, 1947—. Lecturer in Psychology, Columbia University, 1940—. Author of psychological tests including the *Stenographic Aptitude Test*, the *Stenographic Proficiency Test*, and the *Test of Mechanical Comprehension*. Co-author of the *Differential Aptitude Series*. Fellow, American Psychological Association. President, Division of Industrial and Business Psychology of the A.P.A. President, New York State Psychological Association. Vice-Chairman, Division of Anthropology and Psychology of the National Research Council.

Andrew L. Comrey—M.A., University of Southern California, 1947. Served with the U. S. Navy, 1943-1946. Teaching Assistant and Research Assistant, University of Southern California, 1946-1948. Member, Kappa Phi, Psi Chi.

Francis J. Di Vesta—M.S., Cornell University, 1946. Teaching Assistant in Psychology, 1945-1946; Research Assistant in the University Testing Service, 1946-1947, Cornell University. Member, Phi Delta Kappa.

Jerome E. Doppelt—M.A., College of the City of New York, 1937. Interviewer, New York State Employment Service, 1938-1939. Member of staff, Graduate Record Examination of the Carnegie Foundation, 1939-1944. During the war with the Atomic Bomb Project, Los Alamos, New Mexico. Assistant Director of the Test Division, the Psychological Corporation. Member, American Psychological Association, Psychometric Society, American Statistical Association.

Allen L. Edwards—Ph.D., Northwestern University, 1940. Assistant in Psychology, Ohio State University, 1937-1938. Assistant in Psychology, Northwestern University, 1938-1940. Instructor in Psychology, University of Akron, 1940-1941. Psychologist, Special Study Group, Military Intelligence, War Department, 1941-1942. Psychologist, Overseas Branch, Office of War Information, 1942-1943. Assistant Professor of Psychology, University of Maryland, 1943-1944. Consultant, War Relocation Authority, 1943-1944. Associate Professor of Psychology, University of Washington, 1944—.

Author of articles in professional journals. Fellow, American Psychological Association. Member, American Statistical Association, Washington State Psychological Association, Western Psychological Association, Advisory Board of the Washington Public Opinion Laboratory. Consulting Editor, *Journal of Abnormal and Social Psychology*.

J. P. Guilford -Ph.D., Cornell University, 1927. Instructor in Psychology, University of Illinois, 1926-1927. Assistant Professor of Psychology, University of Kansas, 1927-1928. Associate Professor and Professor of Psychology, 1938-1940; Director, Bureau of Instructional Research, 1938-1940, University of Nebraska. Professor of Psychology, University of Southern California, 1940-1942. Director, Psychological Research No. 2, Santa Ana Army Air Base, and of Psychological Unit No. 2 Aviation Cadet Center, San Antonio. Chief, Field Research Unit, Army Air Forces Training Command Headquarters, Fort Worth, Texas. Chief, Department of Records and Analysis, Army Air Forces School of Aviation Medicine, Randolph Field, with rank of Colonel, 1942-1946. Professor of Psychology, University of Southern California, 1946-. Fellow, American Association for the Advancement of Science, American Psychological Association. Member, Psychometric Society, Society of Experimental Psychologists, Western Psychological Association, Society of Mathematical Statistics, Southern California Psychological Association.

John P. Hertel - Ph.D., Cornell University, 1938. Assistant Professor in Personnel Administration, 1938-1943. Associate Professor in Personnel Administration, New York State College of Agriculture, Cornell University, 1943.

Joseph C. Heston -Ph.D., Ohio State University, 1941. Science Instructor and Director of Remedial Work, West Jefferson, Ohio, High School, 1932-1939. Assistant in Psychology, Ohio State University, 1939-1941. Instructor in Psychology, 1941-1942; Assistant Professor of Psychology, 1942-1946; Director of Bureau of Testing and Research, 1944-; Associate Professor of Psychology, 1946-, De Pauw University. Research Consultant, Farm Security Administration, 1940-1942. Author of articles in *Educational Research Bulletin*, *Journal of Educational Psychology*, and *Journal of Applied Psychology*. Member, American Psychological Association, American College Personnel Association, American Association of University Professors, Sigma Xi.

Horacio J. A. Rimoldi -M.D., University of Buenos Aires, 1938. British Council Scholar, London and Oxford, 1939-1941. Professor of Gral. Biology and of Experimental Psychology, 1941-1946; Director of "Instituto de Psicología Experimental", 1946-1947. University of Cuyo, Mendoza, Argentina. Director of "Publicaciones del Instituto de Psicología Experimental," 1942-1946. Frank's Fund

Fellow, University of Chicago, 1946-1947 Guggenheim Fellow, 1947-1948. Author of articles and monographs on mental testing, psychometrics, and general psychology Member, Psychometric Society, Sigma Xi, Psi Chi.

Edward C. Roeber—Ph.D., Northwestern University, 1942. Teacher and counselor, 1935-1943. Professor of Education, Hamline University, 1943-1945. Director of Guidance Bureau and Associate Professor of Education, Kansas State Teachers College, 1945-1947. Assistant Professor of Education, 1947-1948; Associate Professor of Education, 1948—University of Missouri. Author of articles on guidance services. Co-author of a test. Member, American Educational Research Association, American Statistical Association, National Vocational Guidance Association, American College Personnel Association, Psi Chi, Phi Delta Kappa.

Julius Seeman—Ph.D., University of Minnesota, 1948. Teacher, Baltimore public schools, 1934-1943. Graduate Fellow, University of Maryland, 1943-1944. Instructor, University of Minnesota Laboratory School, 1944-1945. Counselor, University of Minnesota Counseling Bureau, 1945-1947. Instructor in Human Development, and Counselor, 1947; Member of Counseling Center and Dept. of Psychology, 1948—, University of Chicago. Associate Member, American Psychological Association.

Merle W. Tate—Ed D., Harvard University, 1947. Research and Teaching Fellow, Harvard University, 1945-1946. Superintendent of Schools, Gooding, Idaho, 1935-1944. Associate Professor of Education, Hamilton College, 1946—. Special Consultant, N. Y. State Dept. of Education, 1947-1948

E. G. Williamson—Ph D., University of Minnesota, 1931. First Director, University Testing Bureau, 1932-1938. Coordinator of Student Personnel Service, 1938-1941. Dean of Students and Professor of Psychology, 1941. Chairman, Committee on Student Personnel Work, American Council on Education, 1944. Chairman, Advisory Committee U. S. Armed Forces Institute, 1942-1946. Chairman, Committee on Disabled Veterans, American Council on Education, 1947—. Member, Scientific Mission of Psychologists to study manpower utilization in Germany, U. S. Department of Commerce, 1946. Author of books and articles on various aspects of student personnel work. Co-Author with John Foley of *Disciplinary Counseling*, in press. President, Division of Guidance and Personnel Psychologists, 1947. President, American College Personnel Association, 1941—.

EDUCATIONAL and
PSYCHOLOGICAL



MEASUREMENT

VOLUME EIGHT, NUMBER THREE, AUTUMN, 1948

TABLE of CONTENTS¹

PROCEEDINGS

AMERICAN COLLEGE PERSONNEL ASSOCIATION

EDITOR'S FOREWORD	409
AMERICAN COLLEGE PERSONNEL ASSOCIATION, DESCRIPTIVE STATEMENT, OFFICERS	410

¹ACPA PROGRAM COMMITTEE

DONALD E. SUPER, Chairman
Associate Professor of Education, Teachers College, Columbia University
WILLARD W. BLAESSER
Dean of Students, State College of Washington
NEAL E. DROUGHT
Personnel Administrator, R. C. A. Victor Corporation, Harrison, N. J.
WENDELL S. DYSINGER
Dean, MacMurray College
JOSEPH S. KOPAS

Director, Department of Development, Fenn College

ACPA CONVENTION COMMITTEE CHAIRMAN

A. C. VAN DUSEN, Associate Professor of Psychology and Director, University College Bureau of Counseling, Northwestern University: *Local Arrangements*.
PAUL C. GREENE, Director, Student Personnel Bureau, University of Illinois (Chicago): *Publicity*.
DOUGLAS BLOCKSMA, Administrative Coordinator, Counseling Center, University of Chicago: *Meals*.
WANDA GUM, Assistant Professor of Social Sciences and Counselor, Student Personnel Bureau, University of Illinois (Chicago): *Hospitality*.
ROBERT WOELLNER, Secretary, Board of Vocational Guidance and Placement, University of Chicago: *Information*.
WILLIAM E. SCOTT, Assistant Dean of Students, University of Chicago: *Posters*

MARCH 29

BANQUET 7:00 P. M.

Presiding, FORREST H. KIRKPATRICK, Dean of Students,
Bethany College

Presidential Address: *The Greatest Tragedy in College
Personnel Work*

C. GILBERT WRENN, Professor of Educational Psy-
chology, University of Minnesota, and President,
ACPA 412

GENERAL SESSION 10:00 A. M.

Presiding, C. GILBERT WRENN, Professor of Educational
Psychology, University of Minnesota, and Presi-
dent, ACPA

The Essentials of a Student Personnel Program

E. H. HOPKINS, Vice-President, State College of
Washington 430

GENERAL SESSION 2:00 P. M.

Presiding, WILMA DONAHUE, Director, Bureau of Psy-
chological Services, University of Michigan

*Professional Standards and Training for College Personnel
Workers*

GORDON V. ANDERSON, Assistant Director, Testing
and Guidance Bureau, University of Texas 451

SOCIAL HOUR

Chairman, WANDA N. GUM, Assistant Professor of Social
Sciences and Counselor, Student Personnel Bureau,
University of Illinois (Chicago)

MARCH 31

GENERAL SESSION 9:00 A. M. (arranged by
NADW²)

Presiding, HELEN RUSH, Dean of Women, University
of Pittsburgh

Minority Groups in our Schools

² National Association of Deans of Women.

EDWARD J. SPARLING, President, Roosevelt College. 460
Who Should Attend our Colleges

HILDA DAVIS, Dean of Women, Talladega College. 473

SECTIONAL MEETINGS

OCCUPATIONAL INFORMATION (with NVGA^a)

1:30 P. M.

Presiding, ROBERT HOPPOCK, Professor of Education,
 New York University

Putting Occupational Information Across

ARTHUR H. BRAYFIELD, Dean of Student Personnel,
 Long Beach City College 485

INTERNATIONAL RELATIONS (Arranged by NVGA) 1:30 P. M.

Presiding, MARGARET E. BENNETT, Director of
 Guidance, Public Schools, Pasadena

UNESCO and Educational Reconstruction

HAROLD E. SNYDER, Director, Commission on Inter-
 national Educational Reconstruction, Washing-
 ton, D. C. 496

FOREIGN STUDENTS (with NVGA) 3:00 P.M.

Presiding, HAROLD W. BAILEY, Associate Dean of Arts
 and Sciences, University of Illinois (Chicago)

Counseling Students from Overseas

CLARENCE LINTON, Advisor to Students from Other
 Lands, Teachers College, Columbia University. . . 501

STUDENT ACTIVITIES (with NADW)

3:45 P. M.

Presiding, JOHN L. BERGSTRESSER, Assistant Dean of
 Students, University of Chicago

Democracy Through Student Government

WILLIAM WELSH, President, National Student Asso-
 ciation and Student, Berea College 522

GENERAL SESSION (with NADW and NVGA)

8:30 P. M.

Presiding, ANNA L. ROSE HAWKES, Dean of
 Students, Mills College

^a National Vocational Guidance Association.

Human Nature and Education

HAROLD TAYLOR, President, Sarah Lawrence
College. 530

APRIL 1

GENERAL SESSION 9:00 A. M.

Presiding, DANIEL D. FEDER, Dean of Students,
University of Denver
*Some Implications of Client-Centered Counseling for
College Personnel Work*
CARL ROGERS, Professor of Psychology,
University of Chicago 540

GENERAL SESSION (with NADW) 11:00 A. M.

Presiding, EDMUND G. WILLIAMSON, Dean of
Students, University of Minnesota
The Therapeutic Handling of Discipline
ROBERT J. MINER, Director of Student Affairs,
Miami University 550

EDITOR'S FOREWORD

THE *Proceedings* of the American College Personnel Association contain the principal papers presented at its annual meeting in Chicago, March 29-April 1, 1948. Due to lack of space, no attempt was made to include the comments of discussants and panel members on the program. The proceedings of the business meetings, a membership directory and the new constitution are being printed separately and will be distributed to the ACPA membership by the secretary of the Association.

The ACPA and other constituent organizations of the Council of Guidance and Personnel Associations held their annual meetings concurrently. March 30 was set aside by all organizations for a joint "Council Day" program. Daniel D. Feder, Dean of Students, University of Denver, and President of the Council of Guidance and Personnel Associations, entitled his presidential address "Next Steps in the Personnel Profession." Other principal Council Day speakers and the titles of their addresses were: Herold C. Hunt, General Superintendent of Schools, Chicago, Illinois, "Our Professional Horizons"; Ben M. Cherrington, Member of the United States Commission on UNESCO and Director of the Social Science Foundation, University of Denver, "The Role of the Counselor in World Peace and Organization"; Sally Butler, President, National Federation of Business and Professional Women's Clubs, Inc., "Jobs, Citizenship and World Understanding"; Leonard D. White, Chairman, Department of Political Science, University of Chicago, "Professional Attitudes among Public Personnel Workers"; and Corinne V. Loomis, President, International Association of Altrusa Clubs, Inc., "Your Sixth Sense." Due to limited space, Council Day presentations are not included in the ACPA *Proceedings*.

The attendance at the 1948 Chicago Convention set a new record, which is somewhat indicative of the growth of professional interest in student personnel. The program emphasized *Personnel Work as a Profession*, and the contributions recorded here will be useful to those people engaged in or interested in all phases of college student personnel service and its administration.

A. C. VAN DUSEN

AMERICAN COLLEGE PERSONNEL ASSOCIATION

DESCRIPTIVE STATEMENT, OFFICERS

THE American College Personnel Association was organized in 1924 as the National Association of Personnel and Placement Officers, but changed its title to the present one in 1931. Its present charter and broad scope in personnel work date from 1937.

Its purposes are to provide for the cooperative association of those persons engaged in college student personnel service and its administration; for the promotion of national and regional conferences to discuss the problems, progress and possibilities of such service; for the formulation and maintenance of professional standards among workers; for the promotion of research and experimentation with the exchange and discussion of reports of the results thereof; for the dissemination among all educational workers of the student personnel point of view as described in the charter of this Association.

1947-48 OFFICERS

C. GILBERT WRENN, President

Professor of Educational Psychology, University of Minnesota

WILLARD W. BLAESSER, Vice-President

Dean of Students, State College of Washington

THELMA MILLS, Secretary

Director, Student Affairs for Women, University of Missouri

FRANCES CAMP, Treasurer

Director, Bureau of Appointments, State University of Iowa

EXECUTIVE COUNCIL

DAN D. FEDER,

Dean of Students, Denver University

GWENDOLYN SCHNEIDLER DICKSON,

Washington, D. C.

DONALD J. SHANK,

Executive Secretary, Institute of International Education

DEWEY STUIT,

Dean of Students, University of Iowa

EDMUND G. WILLIAMSON,

Dean of Students, University of Minnesota

AMERICAN COLLEGE PERSONNEL ASSOCIATION 411

1948-49 OFFICERS

C. GILBERT WRENN, President

Professor of Educational Psychology, University of Minnesota

THELMA MILLS, Vice-President

Director, Student Affairs for Women, University of Missouri

ANNA L. ROSE HAWKES, Secretary

Dean of Students, Mills College

FRANCES CAMP, Treasurer

Bureau of Appointments, State University of Iowa

EXECUTIVE COUNCIL

HUGH BELL,

Professor of Psychology, Chico State College

FREDERIC KUDER,

Professor of Psychology, Duke University

ESTHER LLOYD-JONES,

Professor of Education, Teachers College, Columbia University

FORREST KIRKPATRICK,

Dean of Students, Bethany College

RUTH STRANG,

Professor of Education, Teachers College, Columbia University

Presidential Address

THE GREATEST TRAGEDY IN COLLEGE PERSONNEL WORK

C. GILBERT WRENN

Professor of Educational Psychology, University of Minnesota

It is not easy for me to be humorous in a manuscript. I tried humor on this group several years ago at St. Louis. A parable was told about the characters Personnel, his mother, Mrs. Academic Work, a Doctor Statistic, and even Ferdinand the Bull was brought in. I thought it was clever and humorous, but I was the only one who laughed! So I must perforce be serious and this year I thought I'd go the whole way and be tragic.

Before I put on my Hamlet robes, allow me to be a prologue reader and relate some of the cheerful things about developments in college personnel work. *The assumption of responsibility for personnel functions is more widespread.* Not everyone who talks about the need for student personnel work knows what he is asking for but the use of personnel procedures in the military and in industry during the war has made the term respectable. It is less often confused in spelling with the word "personal." This sounds as though I were trying to be clever and a little cynical. Not so. There is danger, of course, in a term being bandied about too long, and acquiring an over-load of meanings. There is marked advantage, on the other hand, in not having to define and defend the general significance of a function each time you introduce the subject. Compare your experiences of the past year in this connection with those of ten years ago.

I could cheerfully spend some time on the subject of terms, but, then, I am not supposed to be cheerful tonight. I could tell you how I would gladly bury the word "guidance" entirely and use only the term, "personnel" with the appropriate adjec-

tive such as *student* personnel, *industrial* personnel, etc., in front of it. We could then speak in common terms with public schools and non-school agencies of "personnel functions," "personnel administration," "personnel research" and other concepts that are written into the Charter of this Association. I'd also relegate "counseling" to its appropriate position as one of a *number* of personnel functions and not have it substituted in people's thinking for the entire personnel program.

The demand for student personnel workers has been and is acute. This is a sign of growth and a cause for cheerfulness. At almost any one time during the past year I could have named ten or fifteen places where they were looking for an individual to fill some type of personnel job. Each of you who is training personnel workers could testify similarly. This demand and the relationship of our association to it is being recognized in our proposed Directory Information Service. Such a demand has its serious dangers. The present supply of qualified people is not adequate. Furthermore, when trained workers are appointed they are not always given responsibility commensurate with their qualifications as professional people. These dangers are too obvious to need analysis here, but they cannot be forgotten.

There has been progress in the technical tools and procedures made available to personnel people. Objective and projective appraisal tools, individual and group therapy procedures, occupational analysis, mental hygiene procedures in discipline and adjustment are illustrations of technical developments. There is less of an authoritarian, paternalistic flavor to our relationships with students and more emphasis upon individual initiative and responsibility. The housing and social environment of students is taken seriously with the college assuming more responsibility for environment. More could be said. There is healthy growth, even healthy friction. But where are we *failing* most seriously? Professional people may worry too much about themselves but where worry exists I have hope. I am far more afraid of complacency and crystallization. So let us worry a little about our field of work. *At what critical points should we examine the student personnel program in each institution?* There are four such points, the last of which I

believe to be most pervasive and therefore the "greatest tragedy."

1. *Job and Training Standards*

The first of three minor tragedies in the present day student personnel field is the lack of commonly accepted standards of performance and of professional preparation. Had the Board of Representatives of the Council (of Guidance and Personnel Associations) not chosen as its theme "The Status of Personnel Work as a Profession," I would have taken this subject for my address to you. It is my personal belief that college personnel work is not a profession upon several counts, the chief of which is its immaturity in this matter of standards. Some of you were present at the University of Minnesota 25th Anniversary Institute last November when Dean Darley and I analyzed the status of student personnel work as a profession. Between us we applied eight criteria of professional status to this field. They were as follows:

1. The application of standards of selection and training
2. The definition of job titles and functions
3. The possession of a body of specialized knowledge and skills
4. The development of a professional consciousness and of professional groups
5. The self-imposition of standards of admission and performances
6. The legal recognition of the vocation
7. The development of a code of ethics
8. The performance of a socially needed function

We concluded that only in the possession of a body of specialized knowledges and skills had we achieved reasonable maturity. In the other seven we are merely on the way. By some criteria, such as job definitions and standards, and legal recognition, we are only a few inches from the starting post.

Practically speaking, anyone who is persuasive and knows a sufficient number of the right people can be appointed to a personnel position on the campus. There are no well-accepted standards of preparation, academic or otherwise, no licensing, no requirement of internship or experience under supervision.

A good share of this indefiniteness exists because job definitions have not been crystallized. The administrator cannot hire a personnel worker with anything like the definiteness that he hires a stenographer or a professor of mechanical engineering. Part of this is our fault. We have failed to advise the administrator what established people in the field consider job functions and adequate qualifications. One of the previously mentioned criteria of a profession is that of "the self-imposition of standards of admission to the profession and of performance in it."

Independent of the report of our Standards Committee or of later action by the Association I would like to go on record with three recommendations. (These refer only to college and university student personnel work.)

1. Job descriptions and specifications for the several major student personnel functions must be established by the professional workers most concerned. We will meet with cooperation from general administrators in higher education there if the approach is reasonable, but, meet it or not, those with both a vocational and an ethical stake in the field must take the step. If such clarification goes by default to the President or to the Faculty, or to the slow laborious march of time, we have no one but ourselves to blame.

2. Two years of graduate training with the equivalent of one year of full-time experience in college teaching or personnel work should be the minimum training and experience to qualify a person for any college personnel position except those which are definitely subordinate, routine technical, or clerical in nature. This will come. We should anticipate it. The most desirable training is basically psychological but it must include preparation in the field of higher education organization, curriculum, and philosophy, and preparation in the field of social organization and dynamics.

3. Responsibility for the selection of people to be trained, selection in terms of personal qualities and intellectual ability, should be placed squarely upon the graduate school that gives the training.

The condition of vagueness as to job qualifications and training that now exists need not be a minor tragedy for long, but it certainly is so at present.

II. *Initiation and Direction of Program from Above*

The second point of weakness on all too many campuses is that *students and faculty, who have the most to gain from student personnel work, have the least to say about its development and emphasis*. Faculty representation on administrative and policy committees is widespread but not complete. Two summers ago, in talking to the Faculty at the University of Colorado, I said that three groups make up a campus community—the Administration and staff, the Faculty, and the students; and that they should all be represented on policy committees. It was further said that we had won the fight for faculty representation on policy committees but that the student group was still shut out. After this lecture, a visiting faculty member from a large west coast university got up and stated publicly that they had *not* won the fight as yet at his institution. *There was no faculty senate or other legislative body in his institution* (and, of course, no student representation.) On the other hand, the Student Affairs Committee of the Faculty Senate at one large university consists of six faculty members and administrators and six students. This committee establishes policy for all student organizations and for all campus activities in which students play a major part. This is a step in the right direction. Other campuses have gone still further. I recall that, after having commented on this point at the National Conference on Higher Education at Chicago last April, a representative from Wilberforce University in Ohio spoke to me about his campus. He stated that there is student representation on every policy committee in his college, including the administrative committee. Of course, Antioch College has community representation *par excellence* and there are a number of similar situations, but the picture as a whole is not too bright. How is it on *your* campus?

This is so important for the student personnel program because student personnel services are provided to meet presumed student need. Who says the students want or need a given service? When policy is formulated or operation modified *to better meet student need*, who says that it will better conditions or that students will utilize the service? Most often it is an

administrator who has to act. Sometimes a faculty committee will issue the *pronunciamento*. This is too limited and too subject to serious error. Students and faculty must be fully represented on personnel policy committees because fewer mistakes will be made than if decisions are made by the administration alone. And this is no disparagement of honest, hard-working deans. It is a recognition of the fact that all three groups in the college community will see the same need or policy from different points of view. All are affected by the policy also.

A second reason for student and faculty representation is the improved reception under these conditions of any change in policy or program. Every campus should examine itself on these two related points: *Is there a student personnel policy and who made it?* Mr. Hopkins clearly pointed up this same principle in his paper. Administrative machinery must be established that will insure faculty and student contribution to personnel and other policies. Of course the machinery must be present to keep the policy fluid as conditions change or new knowledge becomes available. Beyond this there must be a clear understanding of the difference between policy making, a legislative function, and the execution of policy, an administrative function.

III. *Coordination Through Appropriate Line and Staff Organization*

This leads to the third critical point. *Poor coordination of a student personnel program is frequently the result of an incompletely formulated line and staff organization.*

The organization of personnel work is like the organization of any other function within an agency or institution. That is, it is a matter of line and staff relationships. In this respect, it appears that some would like to avoid facing the question of centralization and direct line authority by talking rather glibly of the *coordination* of personnel work. I am not sure that this is an adequate solution. It is true that coordination of what may be scattered and discrete personnel services must be effected. But this does not dispose of the fact that someone must be *responsible* for the total set of functions involved in a

personnel program. Coordination rather than administration has been proposed in an attempt to avoid offending or frightening certain people. There is always the possibility, however, that it may be *necessary* to change functions or to make individuals directly responsible to someone other than the president.

The individual who directs the program of personnel services is most commonly called a Dean of Students, although he may be called Director of Student Personnel or a Vice President in Charge of Student Personnel. In any event, he has an administrative responsibility parallel to that of the Dean of the College and the business manager or Comptroller. At the National Conference on Higher Education previously mentioned, there was a rather general agreement upon three or possibly four major areas of administration in colleges and universities: the administration of instructional functions, the administration of student personnel functions, and the administration of the business functions. Some would add a fourth, the administration of public relations. In any event, there was agreement that student personnel functions were to be administered by a man responsible for the total program directly to the President. This is a line relationship from each service through the Dean to the President. Vice President Hopkins would interpose an intermediate echelon; a vice president responsible for the administration and coordination of both instruction and personnel. He is doing this at Washington State College.

The second problem, however, is that of staff relationships. Many functions such as that of counseling, are performed by individuals who themselves are not in a line relationship to the Dean of Students. They are responsible to the business manager or to the Dean of the College. Here is where it becomes necessary to see that this function is properly carried out through a staff relationship established between the instructor, for example, and the Dean of Students. By agreement between the Dean of the College and the Dean of Student Personnel, the teaching staff member can be held responsible for the performance of a counseling function as a normal part of his teaching work or where he is acting as a faculty-counselor. Once it is clearly understood that he has only a staff relation-

ship to the Director of Personnel and a line relationship to the Dean of the College, much of any uncertainty as to responsibility is dissipated. The same is true, for example, of those who are responsible for the management of dormitories. The dormitory director may be under the Comptroller of the institution for the strictly business management of the dormitory, but responsible to the Dean of Students for the social and counseling program. He has a line relationship to one and a staff relationship to the other and someone must clarify *which is which* if friction and lowered efficiency are not to result. It might be better still in the larger dormitories to maintain a strict line relationship with a business manager responsible to the comptroller and a director of a program responsible for that function to the Dean of Students. Both arrangements would be by agreement between the Dean of Students and the Comptroller who have a staff relationship to each other.

During the past two years I have spoken or acted as personnel consultant on twenty-six college and university campuses from Massachusetts and Virginia to California and Washington, and this question of *coordination through appropriate organization* has been the most constant major problem. Institutions like the University of Denver, the University of Colorado, and Washington State College have taken great strides toward the establishment of a completely coordinated program. This is a coordination based upon a formulation of policy and careful planning of line and staff organization that has extended over many months or even years. (It is probably unfair to name any institutions in this connection. These are given because their work is fairly recent.) The line and staff problem is acute because *efficiency* and *results* are so vitally affected by good or poor organization of existing services. And the day is coming when a personnel program will be judged by its results, not by paper organization or the indorsements of those in high places.

IV. *Isolation from Significant Influences in the Life of the Student*

What has been said thus far is not too new. What I want to say now is less well formulated and harder to express. Please allow me to blunder ahead and state it this way: *A student per-*

sonnel program on a campus tends to be isolated from four important influences in the life of the student, (a) the home, (b) secondary schools, (c) the college classroom, and (d) the spiritual resources of the campus.

The Home. I recognize that the student of college age is in the process of emancipating himself from dependence upon parents and the home. In spite of that fact the home is the most single important clue to what has happened to him and what is still happening to him. Colleges, however, are not in a favored position as far as contacts with the home are concerned. On many campuses half or more of the students have their homes at some geographic distance from the campus. We secure very little information *from* the home and very little about the home except as the student may give it to us on a personal history blank or in the interview.

Many counselors are appropriately fearful of any direct contact with parents because of the danger of resentment from the student or the danger of a transfer of emotionally immature attitudes from parents to counselor. These dangers are real but no more so than the fact that, although the counselor or the college does not *wish* to act *in loco parentis*, it is frequently forced to do so without adequate knowledge of the parents that it is temporarily replacing.

Any analysis of the significant areas of information about the student will place information on *home conditions* and *parental influences* high on the list. There is little done, however, to establish channels through which such information can come except as action must be taken in an emergency. I have no pat solution that will apply to all situations. There probably is none, but I ask for more sharing of procedures that have worked in given situations and more formal publication on this topic for *college workers*. *The college student* is not so old that we can leave parents and home out of the picture.

Secondary Schools.—We are much too isolated from the secondary schools which give us our students but here there are some hopeful signs of solution. The term "articulation" is familiar to all of us in education. It is written into our Association Charter under the topic of coordination with pre- and post-college agencies. Much that has been done on articulation

has been in terms of admissions standards and curricula. These are important but important also is a flow of information from the high-school counselor to the college personnel worker. This continuity between counselors, between total personnel programs in high school and in college is a weak spot.

I spoke on this subject earlier this month at a meeting in Portland, Oregon, of the Northwest College Personnel Association and four other personnel groups that it had brought together. I mentioned the information form used by the Ohio College Association and a blank developed some years ago by the California Educational Research Association. Dean Golda Wickham, of the University of Oregon, showed me a blank used by the college and high-school deans of women in that area. All of these bring pertinent information *to* the college counselor *from* the high-school counselor.

This same northwest area has an Inter-State Committee on High School-College Relations with a *Sub Committee on State-wide Evaluation of High-School Students*. May I quote two paragraphs from a recent report of this committee, of which E. H. Hopkins is chairman:

After considerable study and deliberation by the subcommittee, the subcommittee recommends that the original concept of the proposal be revised to include a much broader objective than merely statewide testing. It is felt that the program should aim to develop many and various means of individual student analysis (tests, inventories, personal data sheets, cumulative records, refinement of uniform application blanks for admission to colleges and universities, et cetera). In addition, it is felt that the program should not only be limited to methods of individual analysis. It should assume some responsibility for providing professionally sound assistance to high schools in inaugurating the program and in developing sound evaluation and guidance programs generally.

The long range program should be aimed toward better guidance of *all* high-school students, not just those who are primarily interested in going to college. It can be assumed that these objectives will be greatly facilitated by a cooperative and coordinated program. This program should be designed to secure adequate and valid personal data regarding each individual student and to utilize these data to the best advantage of the individual student.

Such a program is articulation on a cooperative basis that will bring the colleges closer to the high-school environment.

It will also increase the mutual respect of each group for the other.

We must not restrict our contacts with the high school to admission data or information blanks. The personnel staffs of the two units must share ideas and problems. This will give us at the college level "the feel" of the high school and its professional problems. This can best be done in statewide or even smaller groupings. *Any college personnel staff* can well afford to initiate professional meetings with the counselors and principals of its nearby feeder schools. Something *can* be done.

The Classroom. On the college campus itself there is isolation of the personnel program from the classroom and from those who teach in it. The teacher and the curriculum are the most important influences in the campus life of the student, important in the sense that he was admitted in terms of curricular requirements and must progress in relation to them if he stays on the campus. We may talk all we like about student interests outside of the classroom, they are manifold and vital, but the classroom puts the "bee" on him and he knows it. He must satisfy the faculty and he knows that. Those of us who believe that the non-academic interests of a student are important should be *most familiar* with curricular and teaching influences in his life because they are so influential in his development. It is a question not only of our knowing what the classroom teacher tries to do but of the teacher knowing *in his language* what we are trying to do. Industrial personnel has gone somewhat further than we have in this interpretation of personnel goals and procedures to the operating staff. They must do this to exist. A former student of mine, Vance Jewson, recently talked to the top management in his company in an attempt to explain the values of careful selection. He used concepts having meaning to them and from his statement I should like to quote a paragraph or two.

1. Why do we have a psychological counseling and testing section?
 - A. Because we are in a competitive labor market in our search for top-notchers.
 - B. Because our primary investment is in people. Every 50 salesmen hired represent \$250,000 investment.

- C. Because of need to reduce training costs and training time.
- D. Because of need for making personnel relationships profitable.
 - 1. Minneapolis Honeywell is not an elementary institution. We have a yen to survive, and we must develop these human relationship programs if we survive. This is not out of a sense of charity but out of a sense of intelligent realization that if our business democracy is to survive more attention must be given to intelligent recruiting, selection and placement of people.
 - 2. Realization that our casual methods of handling employees and our casual methods of interviewing were not profitable. Many of our casual methods in years past have been more comparable to the methods used in judging livestock, or used cars. We have a department head who even now says: "Let me see how the person walks, and I'll tell you what kind of a worker he will be!" Our tendency to form judgments from handwriting, shape of the head, and the physical features, certainly has proven unprofitable. Our personal biases, prejudices, favoritism, unreasonable preference for one group of people over another, general disregard for evidences of promise that may appear on an application blank, or in an interview, with a statement—"It's too bad you don't have the experience"—all are examples of common-sense thinking that lacks scientific evidence to support it. That kind of handling is too costly.
- E. Other factors that have led us to improve our personnel facilities:
 - 1. Unhappy workers (as illustrated by examples presented), supervisors—customers—interfere with production.
 - 2. Hidden costs in correspondence, expense accounts, company cars, errors in judgment.
 - 3. The money cost involved in transfers from one end of the country to another and from one department to another.
 - 4. Our failure to know the psychological job requirements makes for costly errors. All these considerations have led us to feel a need for embarking on a new approach to our problems.

Can we explain *our* services in terms as meaningful to faculty as these are to business management? I think we can but we have not gone far enough in doing so.

There is another approach, that of a sense of cooperative

effort and of joint concern with the teacher. We can help the faculty to help students, once we know the classroom and keep up with curriculum demands so that we can talk a common language with the teacher. Most teachers know the *problems* of dealing with groups and often feel incompetent in handling them except by lecture. Let me give you a paragraph from another former student, David Jenkins, who is with the Center for Group Dynamics, now at M.I.T., soon to move to Michigan. In a paper to a regional guidance conference in New England, Jenkins said:

The counselor, who is the specialist in problems of individual adjustment, should become the resource person for the teachers in matters of making their classrooms good places for adjustment. This would mean that the counselor would not only need to be skilled in techniques of individual contacts with students, but he would also need to become skilled in the techniques of group leadership and group behavior. Solving the problems in the classroom *before* they arise seems much more efficient than having to attempt a cure after the problem has become acute. Preventive counseling seems as sensible as preventive medicine. This suggestion would mean that the classroom group atmosphere and the club atmosphere and the student council atmosphere would all be the concern of the counselor and he would be called upon to advise the teachers and student leaders in ways and means of improving them.

If this sounds idealistic then remember that ideals and new ideas precede intelligent change. To carry out the suggested integration with the classroom we will have to develop skills in group management and therapy and we will have to find the teachers with whom we can work.

Our isolation from the classroom, with marked exceptions of individuals who now teach undergraduate courses or have taught fairly extensively, seems to me to be caused by two factors. In the first place, personnel people are apt to consider themselves as administrators or as psychological specialists and therefore remote from ordinary classroom activities. They forget that the student is *not* remote from these influences and that the personnel worker must take them into account along with counseling and extra-classroom influences. The second factor is the "reverse english" of the first in that the Faculty all too frequently considers personnel workers and the personnel pro-

gram remote from the classroom. This is chargeable to our negligence in permitting the student personnel program to develop without faculty action and support in policy making.

Spiritual Resources.—To state that the student personnel program is isolated from the spiritual resources of the campus and community is perhaps to express the fear that we have over-intellectualized our approach. In a desire to reject sentimentalism and traditionalism we have developed objectivity and a scientific approach. We have not meant to neglect the great intangibles and emotional experiences of life in this process, but it seems to me that we may have been doing so. In our fight against emotionalism we may have "thrown out the baby with the bath water." We have become objective and missed the significance of the subjective realities of experience, become fact-minded and missed the importance of the casual, and become intellectual and missed the dynamics of the spiritual in life.

I am well aware of the difficulty of expressing my convictions on this matter. The above paragraph seems to me to consist of words and more words but to carry none of the intense feeling that I would like to transmit to you. For example, what is meant specifically by spiritual resources? Well, there are two or three answers to that one.

1. *Religion* is one of the great spiritual resources and professional people seem to be afraid of it. Our forefathers were not, and I question the likelihood that we have advanced so far ahead of them in understanding the deeper meanings of life. We have not grown away from religion but religion in reality and in the lives of students may have grown away from us. That is our real tragedy. Personnel workers are aware of the struggle going on within students because of the clash of new ideas and new freedom with what is often an immature and formalized religious experience. Students come to college as immature in spiritual realization as they are in intellectual activity. We help them to greater maturity in thinking. Why not also to greater maturity in living? We are concerned about social life, and health, and accuracy of intellectual conclusions. We have social directors, student organizations, health services and discussion groups to bring about growth in these areas of

living. Why not a program and a coordinator for those experiences which contribute to the *meaning* of life as well as to its process?

Many church affiliated schools have chaplains and chapel services. Even here I have occasionally found the personnel worker aloof from this part of the campus life, he may even be faintly contemptuous of it. There may be good reason to look askance at the formalized religious exercises. There may be much cant and superficiality there. But such a condition does not deny the profound need in the life of a student for a sense of values and relationship to something greater than himself. There are crooked politicians and there is much hypocrisy in politics but our obligation at the polls is all the stronger because of these conditions.

Church affiliated campuses have their own unique problems in the area of religion. They have a great chance but they may muff it. The personnel worker on such a campus should be close to the experiences in the lives of his students that are brought about, for good or for ill, by the religious program of their campus. The state or non-sectarian institution sometimes rises grandly to the occasion, sometimes ignores the whole situation. I am thinking of the chaplains on the staff at Lehigh University and Stanford University. I am also thinking proudly of the recent appointment at the University of Minnesota of a Coordinator of Religious Activities, a man employed by the Minnesota Council of Religion but a member of the personnel staff of the University under the Dean of Students.

2. Another great spiritual resource on the campus is the presence of *great music, lasting art and beauty in many forms*. Beauty is so manifold and is so necessary a part of rich and satisfying living that personnel workers should know well and use well every facility for beauty in the life of the student. By thorough acquaintanceship with the resources for artistic and aesthetic satisfaction on the campus and in the community the personnel worker is himself enriched and thus contributes both directly and indirectly to the emotional life of the student. Professional people of all classes, with personnel workers no exception, are apt to emphasize the intellectual at the expense

of the aesthetic. We must of course "live by our brains" and tolerate no less than our best intellectual effort but the man who sees no more than ideas and facts in life is poor indeed. Does the personnel worker on any campus know his symphony recordings, his local chamber music or a capella choir, the permanent or travelling art exhibitions, the spots from which the best mountain or lake or sunset views can be seen, the wooded walks, the persons in whom greatness lives and to whom to talk is a benediction? If he does not know these and other sources of enrichment how can he understand students and their gropings toward beauty or how can he provide for richness in the daily lives of these students?

3. A third spiritual resource on the campus lies in nothing so pervasive as religion, in nothing so obvious as music or art, but in the *human relationships* that fill our lives. The love of a boy for his mother or Dad, for his friend or for his sweetheart is a powerful influence in his development. This love, or the lack of it, goes far to explain behavior which we say is irrational and emotional. The fact that it is emotional makes it no less real or of no less value. We should not condemn behavior because it can be labeled irrational. We try to understand such behavior in a given individual, we place it in a category and try to get at causes, but do we recognize it as possible good rather than evil and *use* these loves and friendships as resources in the lives of students? Do we do more than joke with him about his sweetheart, or even question his right to fall in love, ignore his loyalty to a friend and shy away from "sentimental" talk about his parents? We are apt to be concerned about a student's grades, vocational choice and methods of study. These are important factors in living but we must at least *appreciate* what the presence or absence of satisfying human relationships does in his life. All we have to do is to see how these influences shape our own lives.

I have been suggesting that the isolation of the personnel program from what I have called spiritual resources is the result of both organizational and personal deficiencies. A program that gives little attention to the factors of religion, beauty and love as resources for growth is a poor thing indeed. Of course

these are sensitive topics, not to be handled crudely or even obviously. But a good program will make it easy for students to develop these phases of living.

As for the personnel worker himself, all that he can do in personal relationships may depend upon his close acquaintanceship with the spiritual resources of the campus. He must be *known* as a person who is more broadly concerned with life than is suggested by his preoccupation with regulations, tests, committee meetings and research studies! These sound pretty dry—as a matter of fact they are! A man's acquaintanceship with musical events, the church, outdoor life, sports, casual and relaxed social life, soon becomes known and students see a man as a man and a woman as a woman, interesting in themselves, not just as a title or the symbol of a regulation.

More than merely being known as a distinct and vivid personality, a personnel worker should *be* one. The man or woman who has become so old or who has dried up his inner resources of emotion so that he no longer enjoys a hearty story, sheds a tear, exhibits "irrational enthusiasms," seeks quiet moments for communion with God, or laughs at his own peculiar behavior, has lost much of his or her effectiveness as a personnel worker and as a person.

This paper has presented four weaknesses in the conduct of our student personnel programs: a lack of qualifications and training standards, poor policy making, poor organization, and isolation of personnel workers and personnel programs from vital influences in the life of the student. These are all tragedies but the last of these is the greatest. It is "the greatest tragedy."

The thesis that I have been trying to present with regard to isolation is that a personnel program whose chief function is the *integration* of a student's experiences and the development of a many-sided life cannot afford to be isolated from the great influences in his life. The tragedy lies in the fact that isolation is the antithesis of integration. We cannot, of all things, remain in isolation if we are to supply the channels for integration. We must know the nature and extent of home influences, we must develop professional relationships and channels of com-

munication with secondary-school workers, we must keep in close touch with college classroom and teacher, and above all we must show in our own lives that we appreciate the warmth and richness of religion, beauty and love. To do all of these things well will challenge us for the rest of our lives. I never want to be satisfied—and as long as I am in the college personnel field I am sure that I never will be.

THE ESSENTIALS OF A STUDENT PERSONNEL PROGRAM¹

F. H. HOPKINS

Vice President, The State College of Washington

Introduction

WITH a strong conviction that today we are confronted with more questions and with fewer answers than at any other time in history, I accepted the invitation to speak on the subject which has just been announced, with the hope that collectively we can find some of the answers to some of the questions, while there is yet time to do something about them.

I am wondering if there is not a fairly common agreement among us with respect to what most of us consider to be the essential elements and qualities of a student personnel program. So far as I know, no modern educator denounces, at least publicly, the "student personnel point of view," which is so well known to all of you, and which I shall discuss briefly a little later on.

At this point, let me state very briefly the outline of my paper. It may be helpful to you in tying it all together: *First*, I shall merely list those essential and specialized student personnel services, which in one form or another, must be provided as an integral and inseparable part of any program of higher education. *Second*, I shall list a group of basic and fundamental principles and assumptions, which are just as essential as the student personnel services themselves. *Third*, I shall make a plea for "optimum" instead of "minimum" essentials in student personnel work, and education generally. *Fourth*, there is a degree of urgency about all of this which we as educators, and as citizens, have not fully appreciated, or, if

¹ Persons whose comments are not recorded, but who gave analysis and recommendations are Robert H. Mathewson, Associate Professor of Education, Harvard University; Lawrence A. Kimpton, Dean of Men, Stanford University; John H. Cornhelsen, Jr., Dean of Men, San Francisco State College.

The session was summarized by Wendell S. Dysinger, Dean, MacMurray College.

we have, we have done little about it. And last, I wish to single out for special consideration and discussion two of the eleven essential principles which I shall propose.

Essential Student-Personnel Services

For the sake of brevity, I shall merely list eleven specialized student personnel services, which I consider to be essential to a sound program of higher education. Each of these could be discussed at length, but let us merely assume that each of these services should be organized, coordinated, and integrated into the total educational program, in a sound and effective manner. With this assumption, I shall eliminate further discussion with respect to the "hows" and "whys" of these eleven essential services. They are as follows:

1. A program of *pre-college counseling, selection, and applicant-centered admissions*.
2. An organized program for diagnosis and counseling of students. This includes both *intensive clinical* counseling and the normal day-to-day educational and personal counseling provided by the faculty and other less professionally trained counselors.
3. An effective *orientation* program, spread throughout the entire first year.
4. *Remedial assistance in various areas* for those students who need it.
5. Definite provision for the supervision, coordination, and integration of the "*co-curricular*" program on the campus.
6. A *student health service*, providing professional services in areas of both physical and mental health.
7. An adequate program of *supervision of living arrangements*, including the food service program. This program must be provided in such a manner as to contribute to the maximum extent possible to the social-educational objectives of the institution, as they relate to the individual student.
8. A well-organized program for administering *financial aids, student employments, post-graduate placements, and job follow-ups*.
9. Special facilities for developing and evaluating the *religious life and interests* of students on the campus.

10. There must be devised and maintained an adequate system of permanent *cumulative personnel records*, which include pertinent information relative to *all* aspects of student life and student accomplishment.

11. At the present, and for the next few years, a special service providing for the *coordination of Veteran's affairs* is an essential part of the total program.

Admittedly, this list of specialized student personnel services is not an exhaustive one. Each of you probably can add others, but generally speaking, these appear to be the *most* essential of the special services included in a comprehensive student personnel program.

Basic Principles and Assumptions

There are certain basic principles and assumptions which are just as essential as are the student personnel services; hence, I should like to call attention to the more important of these:

1. First, of course, is the student personnel point of view, which I have already mentioned. It must be considered as the fundamental and basic principle underlying the total program. You are already familiar with this point of view; hence, as important as it is, I shall not discuss it further. It is well defined and described in the American Council on Education pamphlet, entitled *The Student Personnel Point of View*, published in 1937. In essentially the same form, and in 1938, it was adopted as the first chapter of the charter of this association. It is just as up-to-date in 1948 as when it was adopted in 1938. Certainly, I shall not attempt to improve on that definition or description of the general philosophy of student personnel work. Furthermore, since it is so well described in the charter of this organization, I am obviously aware that I am not telling you anything new.

2. The second principle, to which I have also referred briefly, is the necessity for thorough-going and effective coordination not only between and among the services and principles themselves, but between the services and the instructional program. This coordination must prevail both horizontally and vertically, not only within the institution, but with the pre-college pro-

gram and with post-college adjustments. A program of student personnel services, of and by itself, just does not exist.

3 Education and the processes of education are strictly individual processes. The individual student must be considered as a total unique personality. Consequently, education must be flexible to the needs of individual students. I am reminded of a statement made at this meeting a year ago in Columbus, by Dr. Edwin S. Burdell, Director of the Cooper Union, when he asserted that "if a curriculum does not meet the needs of a given student, it is a *poor* curriculum for him, however much approved it may be by the pedagogical pundits who dreamed it up." This principle of considering the individual student as an individual is absolutely essential to a sound program of education which is based upon the student personnel point of view. Furthermore, there is nothing in this principle which is in conflict with the social objectives of education; there is no dichotomy between education for the needs of the individual and education for the improvement of society.

4 Educational programs, policies, and procedures cannot be established at least in a healthy manner by administrative edict or by fiat. These are matters demanding the combined wits and wisdoms of the entire faculty—with the additional help of the students. In addition, such a democratic policy is psychologically necessary for the success of the program.

5 Another essential principle, at least as it applies in co-educational institutions, has reference to what I consider to be a psychological fallacy apparent in some institutions where separate and parallel organizational structures prevail, one for women students and one for men students. I, of course, have reference to the traditional "Dean of Men" and "Dean of Women" organizations. Therefore, I consider it a sound principle to assume that the problems of men and women students on the campus are of the same order, and that the principles and methods for solving them are the same. Consequently, *one* program of services should be provided, not two. Both men *and* women counselors, teachers, and administrators are needed, and each should work with *both* men and women students—in *one* program.

6. The college campus, both in action and in spirit, must be made an efficient *laboratory for training in democratic living*.

7. While I have no intention, here and now, of listing all of the principles or assumptions basic to education generally, there is another educational or psychological principle which I regard as being so closely akin to the objectives of student personnel services that I should like to mention it. I have reference to the concept of "deferring the selection of a major." The actual process of "selecting a major" is an individual student process, and a mental process. Therefore, there can be *no* institutional edict which says categorically to the student that you may not select a major field of study until "such and such a time." On the other hand, there should be a policy which prevents students from being forced into such decisions too early in their college careers. Along with this principle goes a strong endorsement of the principles and objectives of general education, *but not to the exclusion of training for jobs*. A strong and effective program of pre-college counseling, and high school-college articulation, will eliminate some of the reasons for a deferred-major policy, but it will not eliminate the need for *general education*.

8. As a principle, student discipline, in the broad as well as the narrow sense, must be judiciously administered and in such a manner that the individual student will be strengthened and the welfare of the group preserved.

9. As mentioned before, a student personnel program does not exist, *per se*, in spite of the tendency of most members of the faculty to think so, and to act accordingly. Consequently, when we refer to the essentials of "the student personnel program," we must necessarily refer to "higher education generally." We simply cannot talk about student personnel work, in educational institutions, without talking about *education* itself. Student personnel services *are* education. We must assume that both the objective and the methods of student personnel work, and of education generally, are *absolutely* inseparable, if not identical. Probably you are thinking that it is trite and hackneyed to be told this again—because I am sure that *every* person in this room subscribes to the principle that student personnel work is not merely a fifth wheel—but I

submit to you that in most institutions, even where the essential student personnel services exist as such, the student personnel point of view has *not* permeated the faculty ranks, the administrative staffs, the curriculum committees, the admission boards, the reinstatement committees, nor the educational policies committees which are *responsible* for the educational practices *within* those institutions. The plain truth *is* that student personnel work *still* is considered by far too many faculty members, deans, and presidents simply as a fifth wheel. And they are not referring to the "steering wheel." Consequently, it is imperative that you bring the *entire* faculty, administration, student body, and alumni into active and constructive participation in the practical implementation of the program. I shall say more about this later.

10. In discussing the principle of "faculty participation" and other sound principles of organization, I should like to take issue with one of the currently accepted university organizational patterns, which organizes the President's office into four divisions, namely, the academic, the student personnel, the business management, and public relations. This organization logically calls for a "Dean of Faculty," a "Dean of Students," a "Business Manager or Comptroller" and a "Director of Public Relations or Public Information." This *is* the current trend in the larger institutions. I am aware that on many campuses it has been a *struggle* for status, particularly for the Dean of Students, or Director of Student Affairs. I am indeed happy that on many of these same campuses, the Dean of Students now has a parallel position with the Dean of Faculty. That is appropriate. However, I do not concur with an institutional organization which permits the student personnel program on the one hand, and the academic program on the other, to be so separated that they are brought together *only* by the President. I know of no president who has the time to really integrate these two equally important educational aspects of the program. Consequently, in the larger institutions, I recommend an Educational Vice President, whose *first* and most important job is to *really* bring together, *almost into one*, the instructional program and the student personnel program.

11. There must be a principle, a policy, in fact a plan for

continuous appraisal and evaluation of the program and a *willingness* to adjust and readjust it to fit changing conditions. Closely related to this principle, in fact a part of it, is the compelling need for curricular and student personnel research at all levels of the higher educational ladder, but most important at the post-graduate level, i.e., a follow-up of our graduates, and for that matter of all others who leave our schools before graduation, in order to find out what are the results of our particular brand of education. Just how effectively are we doing what we propose to be doing? This point, also, I shall discuss later.

I have outlined briefly what I consider to be the essential student personnel services in higher education; I have supplemented this list of services with eleven basic or fundamental principles underlying their implementation. There is absolutely nothing new about any of them. These are the services, the facilities, the special bureaus and offices, the specially trained staffs and the basic principles for which we have been striving in our more progressive institutions for the past ten to twenty years. In the majority of our institutions we now have most of these services in one form or another. What more do we want? Where we have had these services, just how adequate and how effective have they been? To what extent have they served their intended purposes, particularly to the masses of the students? And what changes are we going to make, or can we make, if and when the goals of the President's Commission on Higher Education become a reality, that is, with reference to increased enrollments?

Too Much Emphasis on "Minimum Essentials"

Upon a more penetrating reflection about the *real meaning* of student personnel services, the real meaning of education itself, and their common if not identical objectives, I am wondering if we cannot take a fresh and more timely view toward the student personnel program and, in fact, toward higher education generally.

In the first place, I am reminded of the Office of Education publication, published, I believe, in the middle 'thirties, entitled *Minimum Essentials of a Guidance Program*; for our

purposes, it might just as well have read, "Minimum Essentials of a Student Personnel Program." I fear that too many of us, and our deans, our presidents, and our governing boards, think too frequently in terms of *minimum* essentials, when thinking of the student personnel program in our respective institutions. Why should we put ourselves so clearly on the defensive? Why should we show such lack of confidence, such half-heartedness, such weakness of moral courage, when we try to sell the student personnel point of view, and the kind of program which will put that point of view to work? That point of view, with which we are all familiar and to which we all subscribe, is of little consequence until and unless we put it to work in rather concrete and practical situations. If higher education generally is going to face up to the tremendously urgent and complicated social, economic, and political issues in our fast-moving and relatively unguided civilization today, we must *stop thinking* about *minimum essentials*, *minimum facilities*, *minimum qualifications*, *minimum salaries*, *minimum evaluation*, *minimum research*, *minimum standards*, and *minimum services* to the individual and to the community. Indirectly, we have been apologizing for what we have to offer, by asking merely for *minimum essentials*. Perhaps we have not called them that, and in some cases probably we have not even been aware ourselves that we were seeking merely minimum essentials.

If we really *want* to make education effective, if we really *want* to prepare individual students for effective and full living, if we really *want* to save civilization and democracy within the time limits which may be upon us, let us start defining the *optimum essentials* of education, and gear these essentials and their practical implementation to the *immediate* problems facing education today. Let us direct our thinking, and therefore our *actions*, toward the *maximum potentials* from *optimum conditions*, rather than toward the *limited potentials* from *minimum conditions*. Without this kind of forward thinking, without this courage to face reality—and I mean facing reality in 1948, not in the distant future—we might just as well toss whatever strength we have into the military machine and wait for the consequences.

Urgency for More Effective Higher Education

In asking for optimum conditions we must stress the *urgency* of our purpose, and we must also get across the idea that we have no "vested interests," no selfish motives, no axes to grind, but that these conditions are essential, and essential now, if we are to do the kind of job we *must* do—for our own survival.

The recent report of the President's Commission on Higher Education certainly is not based upon *minimum essentials*. It is the most forward-looking, most sensible, and most practical approach to higher education generally that I have yet seen. I am thoroughly convinced that our respective institutional publics will willingly, even *enthusiastically*, support an optimum program for higher education, *provided* we can produce the evidence for its need, *plus* the evidence of our ability to produce tangible results. These two provisos, and particularly the latter one, place a tremendous responsibility upon our own shoulders. I pray to God that we are *big enough* for the job.

The situation is both critical and urgent and further complicated by emotions of fear, distrust, prejudice, selfishness, indifference, and general "uneasiness."

This "uneasiness" and its consequent fearful speculation are very real, and undoubtedly represent a significant "cross-roads" in civilization as we know it. The first question is "Can we survive?" and the second, "How can we adapt ourselves to the new age if we do survive?" In either case, the "cross-roads" unquestionably is here. But let us concentrate on the first question; there is no time in this discussion to concentrate on the second. What consequences are likely, if we become a world frightened by our own disorganized efforts to cope with it? If fear is an individual's greatest enemy, as the psychologists say it is, it is a far greater enemy of a nation—or for a world—because in an individual, therapy is more easily and more effectively administered. We must meet this world-wide fear with *understanding* and with *education*. We must meet distrust, prejudice, selfishness, and indifference, likewise, with *understanding* and with *education*. How else can we bring about changes in the attitudes and motivations of individuals, except through education? But do we have the "stuff" in our-

selves and in our ranks to provide the leadership needed in these critical times? I am confident that we do have, but it is absolutely essential, and essential *now*, that organized education in this country come decisively to grips with the worldwide crisis of mankind. We cannot afford to let this crisis just "take its course."

President Hutchins has made an apt statement, in speaking of college administrators generally, and I am strongly inclined to believe that his statement characterizes both you and me, regardless of our particular jobs. He says, in effect, that we get ourselves so busy with the immediate and urgent problems of our jobs that we never have the time, nor do we take the time, for the *really important matters* needing our attention. I believe rather keenly that we get ourselves so "wrapped up" and so involved in our jobs, and in our daily routines—and many of them *are* routines—that we have not taken time out to provide whatever leadership and attention we are capable of providing in matters related to the really important issues in education and to our national and international situation. We get ourselves so "tied up" with details and seemingly immediate and urgent matters, that we fail to view our jobs, our institutions, our students, and particularly our products, with the kind of perspective needed for objective, or even subjective, evaluation of what it is we are trying to accomplish.

The second urgent and compelling reason for a fresh view and a revitalized program of higher education based upon the student personnel point of view is the fact that too often our graduates simply do not measure up to our expectations, to their own capacities, or to the demands of the society of which they become a part. While there is continuously mounting evidence which clearly indicates this weakness in higher education today, for the sake of brevity I shall use but a few references. Perhaps the overall problem in this connection is nowhere better summarized than in the Editor's Foreword to Robert Pace's report of the Minnesota study, completed just prior to the war, entitled *They Went to College*. I am aware that this study has been widely quoted, but I hope it is *repeatedly* quoted, especially to audiences of educational administrators and college faculties generally. This foreword, as you

would expect, is based upon the findings of the study, in which Pace reported on the differences between college graduates and non-graduates, at the University of Minnesota. I quote:

We need desperately to know why there appears to be little or no difference between graduates and nongraduates, between high-ranking students and low-ranking students, after they have been a decade away from the campus. Why most of them appear to want security and contentment instead of taking a vigorous delight in 'looking upon the bright face of danger' and welcoming blood-stirring change. Why, if we have taught them - far above their fellows - to think critically, they are in after-college years so obviously uncritical and inconsistent in their thinking. Why, if we have taught them to read good books, most of them read only 'slick' magazines of huge circulation, newspapers, a few books of a standard below that of the freshman English class. Why, in a democracy, the most highly educated people we have fail so miserably to engage in community and political activities.

With respect to the personal lives of the graduates and non-graduates, let me quote from the summary:

The General College faculty, however, were led to ask pointedly whether or not colleges should feel satisfied with the picture described here. Their overwhelming conviction was that they should not.

One of the factors giving rise to this dissatisfaction was that in almost none of the many activities and characteristics surveyed were there any differences between the graduates and nongraduates. Those who completed their college programs were no different in any notable respect from those who failed to complete them. What few differences there were seem almost random or accidental rather than expected outcomes of an additional two or three years of college training.

And, from page 78 of the same study, I quote:

If these scattered and limited differences between graduates and nongraduates represent the measure of effective education, then it behooves educators to take more specific, direct, and active responsibility for the life goals, philosophies, personalities, cultural interests, and recreations of the young people with whom they work.

Much more could be said with respect to the *need* for an effective student personnel program, one which is thoroughly meshed into the total educational program and which would be capable of producing products of which we could be proud.

Just take a look at our attrition rate, throughout the country, where from fifty to sixty-five per cent drop out before embarking upon the third year of college. Just take a look, without even a pretense at scientific evaluation, at the appropriateness and effectiveness of the training which from year to year is being given those who do drop out along the way. Just find out, in your own institutions, what percentage of *your* students do not know what occupational objectives they wish to follow, or for what general occupational pattern they are best fitted. Just find out, *in your own institutions*, the extent to which individual students *are* following intelligent and *planned* college careers. Just find out, if you can, on your own campus, how many who would be far better adapted to terminal and more practical training, are still endeavoring to conform to high academic requirements, demanded by prescribed curricula leading to the bachelor's degree.

While this picture is a rather discouraging one, representing a combination of a critical, serious, and urgent national and international situation, and what appears to be an inefficient and ineffective educational system, I do feel that there *are* ways out. My faith is still stronger than my fears, but faith, and faith alone, is of little value if used as a cloak. We must be realistic. This is no time to "play ostrich." With American optimism what it is, and since we have not only the proved capacity for effective "esprit de corps," but also the proved capacity to actually "do things," let us examine the problem from the more positive side. The force of necessity, in America, still remains a powerfully strong psychological weapon, but we cannot stand idly by.

Although I do not recall the specific reference, I recall reading not long ago in another article by President Hutchins in which he made some such statement as this:

As a means to a peaceful world, education is either everything, or else it is nothing; if it really is everything, then it should be encouraged, implemented, financed, and made to work; if on the other hand, it is nothing, there should be no more time nor money wasted on it.

I think we agree on the fundamental importance of education; I think we agree that it *is* everything, so far as progress is con-

cerned. We agree on the wisdom of the student personnel point of view in education, but apparently we have had little agreement on what constitutes the most fruitful means of putting this point of view and education to work.

Two Positive Approaches

In the light of this setting: first a war-torn, tattered, and unstable world; second, a rather unenviable record of civic and social participation and contribution on the part of the average college graduate, as compared with non-college graduates; and, third, a heterogeneous and conglomerate mass of student personnel and educational practices being used and misused throughout our respective institutions; I should like to focus more attention on two of the eleven essential principles which I mentioned at the beginning of this paper: *first, the principle of getting more active faculty participation in the implementation of the student personnel point of view, and secondly, the need for many more facts, i.e., curricular and student personnel research.* While all of the principles and assumptions outlined earlier are important, I feel that the greatest immediate need is for the practical implementation of these two postulates.

Regarding the first: if we are going to have an effective educational program, based truly upon the student personnel point of view, that point of view must be known, believed, accepted, and *practiced* by the teaching faculty, by the administration, and by those responsible for the specialized services within the institution. In your efforts to secure this participation, probably you will succeed in getting ready acceptance of your objectives and purposes, but the translation of these into "action patterns" will be the more difficult hurdle. And yet, we have talked and we have talked and we have talked about these objectives, in high sounding and highly acceptable terminology; but if we really mean what we say, we must devise the means to accomplish this objective. This is essential—but "How"? This is where you come in—in fact, it is *your* first essential. We must first create an organization, a faculty, in fact an institution, which by its composite and coordinated efforts, by its sound educational policies which will have been put into action, will do for the masses what a few counselors and a few instruc-

tors have been trying desperately to do for too few students. Certainly, I do not mean that you will eliminate the need for individual counseling by specially trained and qualified workers—we will always need the specialists and probably more of them than we can expect to get—but we will get far greater results with the masses of our students, if we provide on a wholesale and preventive basis, what we have been attempting to provide on a retail and therapeutic basis. It is comparable with the effects of a public health program. You need relatively fewer physicians in a community having an intelligent and carefully administered public health program than in communities where no such program exists. Such are the possibilities of an educational program when all of its forces are brought to bear upon common objectives.

This would be merely empty “mouthing” if I did not believe that we *can* go a long way, in a relatively short time, toward the development of this kind of a program. You are the people who must carry the torch; you must become crusaders, and maintain that zeal and enthusiasm, along with a necessary amount of patience and diplomacy, until the process begins to “take.” More than ever before, you must devise better means of informing your faculties, and doing it repeatedly, of the significance of the student personnel point of view. In weekly or periodic staff bulletins, report and review individual cases, report and review the kind of facts which Pace reported in the study of the 951 graduates and non-graduates of the University of Minnesota. If possible, find out these facts from your own institution and your own students. Also, report continuously and repeatedly what other institutions are doing, significant, urgent, and compelling statistics on such matters as student failures, student drop-outs, the criteria for success and the reason for failure on post-graduation jobs, and the failure of college graduates to assume more than a minimum of civic responsibility.

If no such weekly or bi-weekly staff bulletin exists in your institution, *create* one, with the support and participation of the important policy committees of your institution. There is a virtual wealth of current, factual, pertinent and suggestive information available, which, if handed on to our teaching

faculties and administrators. I am positive would stimulate both thought and action.

While heretofore on our own campus we have published some such information in our weekly staff bulletin which is a combination news sheet and campus calendar of events, I have in mind creating a new staff bulletin, to be devoted *exclusively* to matters of educational improvements policies, principles, facts, news, research results, surveys, and opinions which are pertinent to our own educational program. If sufficient and pertinent information could be gathered, carefully screened, edited, or rewritten in a fashion for your own campus consumption, and then disseminated to your entire faculties, over a period of time I think you might be amazed at the effect. Stimulate faculty discussions, faculty meetings, panels, visiting lectureships, radio programs, debates, and seminars on current issues in education. Ask faculty members to review, for publication in the weekly staff bulletin and for small group discussions, such books as Howard Mumford Jones' *Education and World Tragedy*, Donham's book on *Education for Responsible Giving*, Cord Meyer Jr.'s recent book entitled *Peace or Anarchy*, the Harvard Report on *General Education in a Free Society*, the A. C. E. *Cooperative Study in General Education*, or McGrath's recent book, *Toward General Education*. There are literally dozens of such books, pamphlets, surveys and reports which should be called to the attention of your teaching faculties and administrators.

Why not prevail upon your President, your Dean of Faculty, or your Budget Committee for enough money - and it would take a surprisingly small amount--to purchase sufficient copies of the *Report of the President's Commission on Higher Education* so that each department in your college would have copies of the complete report, and insist that it be rotated among the members of the respective departments until all had read it. The reading of that report should be an absolute requirement of every person who justifies his name on a college faculty payroll. Also, for a very small budgetary consideration (75¢ per year per person), your institution can subscribe to *Higher Education*, the semi-monthly publication of the Higher Education Division of the U. S. Office of Education; I recommend it as a good in-

vestment by the institution for every single member of your faculty. If that cannot be done, I recommend it for all Deans, Directors, Departmental Chairmen, and the members of your curriculum and policy committees. Other publications of *significant* value, and of a form adapted for wide and economic circulation, are the *Intercollegiate Press Bulletins*, a weekly news letter devoted to current college events published by the American College Public Relations Association, and *What the Colleges are Doing*, published by Ginn and Company. From these three publications alone, in addition to your own professional journals, and perhaps *School and Society*, you will have enough source material to publish a short but extremely valuable weekly or bi-weekly mimeographed bulletin to your faculty colleagues.

How many of you take time, or make time, to visit *regularly* with members of your instructional staff to discuss with them common policies and procedures, institutional objectives and your *joint* responsibilities for achieving them, to encourage or solicit constructive criticism from them, to demonstrate through a discussion and consideration of *their* teaching and research interests that you understand and appreciate *their* important role in the overall educational program of the institution?

This year on our campus we inaugurated a series of monthly faculty meetings, for *all* members of our faculty; such meetings are devoted *exclusively* to lectures, discussions, and panels covering matters of current educational and faculty personnel policies. Although our faculty probably is as busy and overworked as the average faculty, I am happy to say that these meetings have been unusually well-attended and unusually successful. As a result we have set aside \$10,000 for next year, to bring to the campus outstanding authorities, consultants, and visiting lecturers on matters of current educational interest to the faculty

Another example of the degree of genuine interest in, and willingness to participate in and contribute to, a dynamic program of educational improvement on our own college campus, was demonstrated a few weeks ago when I called for volunteer memberships on a large number of subcommittees of the Educational Policies Committee. It was understood that

these subcommittees were to delve *deeply* into such matters as "improvement of instruction and the relating of instruction to contemporary issues," "integrated courses," "instruction in community, national, and international affairs," "student counseling," "graduation requirements," "a deferred-major plan," "the integration of student activities into the educative processes," "terminal education," "how to teach worker education," and a number of others. With faculty members already "committied to death," you would be astonished to know that we had approximately one hundred and fifty faculty volunteers who indicated both an interest in, and a willingness to serve on, such committees. Those committees are going to stir up a lot of grass-roots thinking on the Washington State College campus within the next few months, and the next year. Our faculty has just completed working out its own plan for faculty evaluation. This means appraisal of faculty performance and effectiveness in areas of instruction, research and publications, and in over-all benefit to the institution and to the community. They have now asked that the faculty be evaluated by the students and by the alumni. They *are* deeply interested in self-improvement and in the improvement of instruction and education generally. Our faculty voted, as a result of their own deliberations, not merely to adopt a deferred-major plan, but to place the administrative responsibility for the entire lower division in the hands of the Dean of Students and his student personnel staff. This plan has been in operation for two years. They are sold on it, and are improving it.

In connection with institutional committees, may I suggest also that you, as student personnel workers, have an important and urgent role to play in the determination and implementation of educational policies within your respective institutions. Therefore, you should see to it somehow that you and other selected members of your student personnel staffs are placed on such committees as the curriculum committee, educational policies committee, admission policies committee, and other important committees charged with policy determination functions. Also, on your important policy committees, place some of the best students on your campus. If you have not already tried this, I think you may be both amazed and pleased

with the nature of the contribution which they can and will make.

The second postulate on which I wish to focus more attention has to do with the dire need for *curricular and student personnel research* to guide us in our policy determinations. How on earth can we expect to develop new curricula, new programs of study, proper combinations of general and specialized courses, integrated courses which we know are effective, adequate vocational and other counseling services, student activity programs which are supposed to train for democratic living, sound admission policies, sound re-instatement policies, sound and defensible graduation requirements, when we do not even know what the effects and the results of our present curricula, instruction, and services are?

We set up admission policies without knowing what really makes for success in college. For example, we are still clinging to high-school graduation and certain patterns of high-school subjects, when we have known for years that there is no scientific basis for believing that high-school subject-matter patterns are related to academic success in college.

Each year we admit large new freshman classes, the largeness of the classes being made possible by the graduation of one group and the dropping out of another. Usually the drop-outs will exceed the graduates; and yet, each year, for the most part, we continue to expose each new class of freshmen to the same curricula, the same requirements, the same rules, and the same services which had already eliminated more than 50 per cent of each preceding freshman class for the past several years. For those who were not eliminated prior to graduation, i.e., those who succeeded in being graduated, we frequently do not even have their names on the alumni mailing list, to say nothing of our failure to know anything about what they are doing, how they are adapting to their jobs or how successful they are technically and professionally. We do not know whether or not they have been able to cope with the complexities of society, whether or not they are leaders in their respective communities, whether they are criminals, or patients of mental hospitals. But—each year, it becomes our first order of business to take care of our new freshmen! That comes first!

Can you imagine any other big business, or even small business, which would continue production year after year, with heavy investments required year after year, without even an inquiry as to the quality and the substantiality of its product? Fortunately, for most manufactured products, the consumers are the primary appraisers and evaluators of the products. To a certain extent, this is true with the consumers of college graduates—the employers; but when the graduate does not happen to measure up, it is the graduate's fault—not the fault of the school. It just could not be the school's fault—after all, look at all the good men who have been graduated from that school! Yes, even in spite of the school! Frankly, I do not really mean to say that all our curricula, and all our instruction, and all our services, are bad; I know they are not. But, I am saying that in most cases we do not know how good nor how bad they really are. We should certainly find out.

In addition, we should know far more about the students who come to our campuses for educational training. Before we admit them, we should know something of their aptitudes, interests, personality characteristics, as well as their scholastic achievements. We have no right to admit them, and then later to tell them that we are sorry, but we just do not have what they need. Usually we do not even tell them that; they struggle along for a year or two, and then discover for themselves that not only are they swimming upstream, but up the wrong stream. We need to know what the individual needs of our prospective students are, and in addition we need to know the overall pattern of individual, social, and employment needs of the regions which we serve, if we are to fulfill our missions efficiently and to the best of our abilities.

In my opinion, the effectiveness of any educational institution, and therefore the effectiveness of its governing board, of its President, and of those others responsible for educational policies and practices within the institution, should be measured by the quality of the product which the institution produces. This, of course, assumes that the qualities and qualifications of the product would have to be measured, and measured consistently and continuously. It means, in all institutions, that such studies as the one reported by Pace at Minnesota should be

made, and the data should be kept current. "But," you say, "this would cost money." I can only say that it would cost far less than what is now being spent providing teaching faculties, services, housing, and other provisions for the students who are "swimming up the wrong streams," not to mention the loss of time, money, prestige, and "face" of each student in that category. Consequently, if we really are interested in developing an educational program based upon the student personnel point of view, let us find out a little more about the individual needs of those students who come to us for training, the social and employment needs of the regions which we serve, and the effects and results of the training which we are now providing for our students. *This research is essential*, and you are the people who are best qualified and who are in the most strategic positions to carry it on. You are the persons to see to it that the findings of such research become the common property of your administration and your faculty.

Summary

- 1 The *essentials of a student personnel program* consist of:
 - a. The student personnel point of view as applied to all educational processes;
 - b. A group of at least eleven specialized student personnel services, requiring specially trained counselors and technicians;
 - c. Certain basic and fundamental principles and assumptions which apply to all phases of the program. Eleven such principles and assumptions have been specifically mentioned.
2. We must strive harder than we have ever striven before for "optimum" conditions and the wherewithal to carry out an effective educational program.
3. We must devise countless practical and spirited means of transforming the student personnel point of view into institutional and *faculty action* patterns—not only as a long range objective, but as an immediate one.
4. Through scientific research we must *get at the facts* which are basic to an intelligent and effective program. *We must know wherein we are ineffective or weak.*

5. There is a degree of urgency which we can ill afford to ignore. In short, all that I have been trying to say is that it is high time that we come down from the ivory towers, that we face the urgency before us, that we follow the courage of our convictions, and develop an educational program which will bring about the full realization and meaning of *democracy in action*. This means an educational program which will turn out emotionally mature men and women *capable, willing, and desirous* of acting intelligently in a world where men depend upon men, and in a civilization which depends upon understandings among men. In the words of the President's Commission on Higher Education, "the responsibility for the development of these personal qualities (in our students) can no longer be left, as heretofore, to some courses, or a few departments, or some scattered extra-curricular organizations; *it must become a part of every phase of college life.*"

•

PROFESSIONAL STANDARDS AND TRAINING FOR COLLEGE PERSONNEL WORKERS¹

GORDON V. ANDERSON

Assistant Director, Testing and Guidance Bureau, University of Texas

THE topic of the present discussion is Standards and Training for College Personnel workers.² The problem is that of arriving at some judgment of what standards should be set for college personnel work, and how those standards can be satisfactorily applied; and also of outlining what should be considered a minimum essential training program for college personnel workers, probably also with recommendations of what training beyond such minima personnel workers should have.

In approaching the general problem, it is probably desirable to review the scope of college personnel work. No disagreement on this is likely. It is generally recognized that college personnel work concerns itself with the extra-instructional or extra-class activities of the college student; but that is not to say that it is unrelated to class work and to the curriculum. This point is often lost sight of, unfortunately, and personnel work is too often thought of in terms of housekeeping and policing functions. We need frequently to remind our academic associates that the effective functioning of a personnel program contributes to more effective instructional work. On the one hand, when students engage in class work, it should be after steps have been taken to insure that they are ready to profit by such instruction, and that *its* objectives are also those of the students. Furthermore, when the student is as-

¹Persons whose comments are not recorded, but who gave an analysis and made recommendations are: Ruth Strang, Professor of Education, Teachers College, Columbia University; Willard W. Blaesser, Dean of Students, Washington State College; Dewey B. Stuit, Dean of Student Personnel Services, University of Iowa, summarized the session.

²This paper embodies the report of the Committee on Standards and Training. The Committee is composed of: Leona E. Tyler, Chairman, Director of Bureau of Personnel Research, University of Oregon; the author; Anna L. Rose Hawkes, Dean of Students, Mills College; E. H. Hopkins, Vice-President, Washington State College, and E. K. Smiley, Director of Admissions, Lehigh University.

sisted in making a more effective adjustment in the total college situation, it may be expected that his general orientation will be such as to make him a more efficient student in his course work. This does not mean mere attention to administrative factors, such as to have students well housed, supplied with adequate health services, to have available a diversified social program, etc. It goes *beyond* this to a consideration of the interpersonal and intrapersonal dynamics which influence student adjustment in even the best of physical facilities. Furthermore, a well-functioning personnel program has values of its own to contribute, so that out-of-class experiences which students encounter themselves also become influences contributing to the growth and maturity of the student. While colleges are institutions oriented around classroom instruction, there are none of us who fail to recognize that students learn just as much outside as in the classroom, and often develop skills and acquire information which a classroom situation would be unable to deal with effectively.

But we cannot arrive at any conclusion concerning standards and training from such broad generalities. We must be more specific in our approach. Jones² has discussed three approaches to the problem of the preparation of personnel workers, which are relevant to the present discussion. These are:

1. The attempt to find what personnel workers do; what their duties and responsibilities are. These attempts have taken the form of job analyses from observation and from the statements of personnel workers themselves.
2. The discovery of the patterns of knowledges, skills, attitudes, interests and characteristics necessary for the successful performance of these duties and responsibilities.
3. The determination of types of courses, training and experiences that are most effective in developing the patterns of abilities necessary for success on the job.

Very little research has been found in any of these approaches. Reliance has been placed rather on statements made by personnel experts of their judgments with regard to each point.

The *whole* problem of "*standards of practice*" in personnel work is pretty well left aside in the present discussion insofar

² Jones, Arthur J. "Preparation of Personnel Workers." *Review of Educational Research*, XV (1945), 185-190

as it relates to what kinds of situations personnel workers should operate in, and to the resources and facilities which should be made available for effective functioning. This might raise the problem of institutional standards. The question of the relationships which should hold between personnel workers and other college officials, and the relations of personnel workers with students are given little attention, although there has been a rising amount of literature on these points and on the closely related one of professional ethics, during the past few years.

The problem of training standards is dealt with rather specifically, and in dealing with the general topic of training, a number of related questions present themselves: (1) Should we concern ourselves with the common training which all personnel workers should be expected to have, or should training for specialties within the field of work be defined as well? (2) Should different standards be set for various types of college personnel positions? (3) What recommendations should be made with respect to the possession of advanced degrees? (4) Should experience requirements be established, including experience in non-academic work? (5) Can standards be set in such a way that persons will be selected for and survive in training programs who possess the personality characteristics generally considered desirable? (6) How should standards which are agreed upon be handled administratively? Should some form of certification be introduced?

The first difficulty that confronts us as we attempt to set up reasonable standards is the wide variety of positions in different institutions. Rather than to attempt a job analysis of every position carrying some sort of personnel title, it seems more feasible to begin with a list of *functions* which it is quite generally agreed that the institutions should perform. More than ten years ago a committee of the American Council on Education formulated a list defining 23 personnel functions. We have used this list, somewhat condensed, as a comprehensive outline for organizing an approach to standards. These functions are as follows:

1. Interpreting institutional objectives and opportunity to prospective students, their parents, and to workers in secondary education.

2. Admitting students in cooperation with secondary schools.
3. Orienting the new student to his college environment, and keeping him continuously and adequately informed of the educational opportunities and services available to him.
4. Providing counseling services which, with the aid of diagnostic facilities and other referral agencies, assist the student in adjusting to and planning for his educational, vocational, emotional, social, and religious growth.
5. Determining the physical and mental health status of the student and providing appropriate health service.
6. Providing and supervising an adequate housing program for students.
7. Providing and evaluating an adequate food service for students.
8. Developing, supervising and evaluating a program for student activities.
9. Assembling and making available information to be used in improvement of instruction and in making the curriculum more adjustable to the needs of individuals.
10. Supervising and coordinating the program of financial aid and part-time employment of students, and counseling the student who needs to obtain such help.
11. Assisting the student to find appropriate employment when he leaves the institution and following up the student after he has left the institution.
12. Keeping student personnel records and making them available to the proper persons.
13. Regulating student conduct to the end that the individual will be strengthened and the welfare of the group advanced.
14. Maintaining student group morale by evaluating, understanding and developing student mores.
15. Carrying on research studies designed to evaluate and improve personnel functions and services.

It is possible to combine these functions in such a way as to represent groups of personnel occupations which apparently require similar types of special skills. We have centered our recommendations on the levels and types of training to be required around this grouping. The *titles* of positions under which some of these duties are carried in some institutions may have been omitted, since no survey of positions was made, but such omissions can easily be corrected.

Since all personnel workers have as their central aims the welfare of the individual student and his adjustment to the college situation both in and out of the classrooms, it has seemed to us that training for *all* should be built around a common core. This should involve *information* with regard to in-

dividuals as individuals and as members of groups. It should also include the development of skill in identifying individual needs and problems and handling interview and group leadership situations constructively. The committee has not specified any certain number of courses or length of time to be spent in covering any of these areas. The Association may eventually wish to set up an examination to determine competency rather than to assume that exposure to courses guarantees it. We have not given specific course names, since at different institutions the same type of training is offered sometimes in the Department of Psychology, sometimes in the College of Education, sometimes in a special personnel training program. We would recommend that in each of the areas listed, the trainee take enough work to acquire more than a superficial familiarity. Basic *habits of thinking* about human beings from a consistent scientific point of view: actual *skill in performance* of fundamental personnel functions are the goals at which we should aim. The common core should include the following:

1. The psychology of personality, its development, organization and dynamics. This should include a number of viewpoints in regard to the dynamics of human adjustment (e.g., Freud, Lewin, Allport, *et al*). Flexibility rather than dogmatism in dealing with individual cases should be promoted. Work in abnormal psychology and mental hygiene should be included in order to familiarize the trainee with the symptoms of serious mental disorders, particularly so that he will know when referral for special therapeutic services is necessary or desirable. The trainee should become familiar with the general course of human development from infancy to maturity, with special emphasis on the characteristics of the adolescent period. He should be able to recognize signs of immaturity for what they are and to understand the characteristic problems of the older adolescent.
2. *Social Psychology* with especial attention to the nature, origin and extent of individual differences, and to group organization and inter-group and intra-group relationships.
3. *Principles of learning*. Since the personnel worker's job is to facilitate the educational process for the student, he needs to have a thorough understanding of how people learn, and what factors aid or disturb the learning process.
4. *Mental tests and their interpretation*. Here especially superficiality is to be avoided, since mental tests are not in a stage of evolution where the results can be taken at face value and

used uncritically in the analysis of an individual personality. Statistical understanding should involve more than the recognition of the meaning of different types of relative score. Emphasis should be placed, however, on information leading to the correct *interpretation* rather than on the administration and scoring of tests or the manipulation of statistical formulae.

5. *Interviewing and counseling procedures.* This training should include actual practice in both the traditional interview techniques designed to obtain and to impart information and the newer non-directive procedures.
6. *Higher education.* Its philosophy, organization and administration.

Besides the course work taken in regular classes, the standards for each personnel position should provide for a period of supervised *experience* in the general type of work the trainee desires to enter. Again no exact number of hours or months has been set in this report but can be added if the Association feels it should be specified. Such experience should not be too narrowly defined. For a person who wishes to be a dean of men, for example, it would not be necessary that he actually take an internship in a dean's office. Work as a dormitory sponsor, student intern in the Counseling Center, etc. would be just as acceptable. The *grouping* of personnel positions forming the basis for the special training recommendations could determine the organization of this part of the program as well. Thus an institution training student personnel workers could set up internships in the record-keeping division, the student activities division, the counseling division, and the employment service.

These two requirements, a common core of studies and a period of supervised experience, would hold for all types of personnel positions. For each of 5 defined types, we have also set up certain special training requirements and a minimum academic degree. They are as follows:

- I. Positions requiring assembling and recording of data about individuals, formulating and carrying out of admissions policies, keeping of records, and planning and carrying out research studies from personnel files.
 - A. Names of positions commonly found in this group:
 1. Director of Admissions.
 2. Registrar.
 3. Personnel Research Director.

- B. Academic training minimum—one year beyond the Master's degree. (PhD recommended for Research Director.)
- C. Special types of training in addition to common core:
 1. Business and/or educational administration, covering office management, record systems and filing, accounting, etc.
 2. Large scale filing and personnel accounting methods, including use of punched card equipment.
 3. Advanced statistical methods.
- II. Positions requiring direction of student activities and organizations, orientation of new students, regulation of student conduct, etc.
 - A. Names of positions commonly found in this group.
 1. Director of Student Activities.
 2. Dean of Students.
 3. Dean of Men.
 4. Dean of Women.
 5. Student Housing Director.
 - B. Academic training minimum—one year beyond the Master's degree, positions 1 to 4; M.A. degree for 5.
 - C. Special types of training in addition to the common core.
 1. Education courses emphasizing extra-curricular activities, diagnosis of learning and motivational problems, development and use of personnel records.
 2. Sociology courses on group interaction, group therapy, etc.
- III. Positions chiefly concerned with the counseling of individuals.
 - A. Names of positions.
 1. Director, Student Counseling Service.
 2. Personal Counselor.
 3. Vocational Counselor.
 4. Educational Counselor.
 - B. Academic training minima
 1. Director—Ph.D in psychology or its equivalent in education or combined program emphasizing counseling.
 2. Personal Counselor—Ph.D in psychology, with specialization in counseling.
 3. Vocational and Educational counselors—Master's degree in psychology, education, or combined program.
 - C. Special types of training.
 1. Occupational analysis and information.
 2. Advanced educational and psychological measurement.
 3. Psychotherapy.

IV. Positions chiefly concerned with testing. (In many institutions, these would be included under III above, since the testing of individuals is a part of the counseling process. Some institutions have, however, a large scale testing program involving function 9, "assembling and making available information to be used in improvement of instruction," as well as the service to individual students themselves.)

A. Names of positions commonly found in this group.

1. Director, Student Testing Service.
2. Psychometrist.
3. Director of Research (listed also under I above).

B. Academic training minima:

1. Director —Ph.D. in psychology or education with emphasis on evaluation and measurement.
2. Psychometrist —Master's degree, with emphasis on administration and interpretation of tests.

C. Special training.

1. Advanced statistical methods.
2. Test construction.

V. Positions concerned with student employment and graduate placement, and administration of financial aids such as loan funds, scholarships, etc.

A. Names of positions.

1. Director, Employment Service.
2. Interviewer.

B. Academic training minima:

1. Director —Master's degree in psychology, business administration, or education, with emphasis on vocational problems.
2. Interviewer —Bachelor's degree in one of the above-mentioned fields.

C. Special training.

1. Occupational analysis and information.
2. Office management, record systems, etc.

Two types of personnel service are included in the list of functions with which we started, but not in the special training recommendations. The first of these consists of positions for which recognized standards are already set by some other accrediting agency. Physicians and nurses in the Health Service, and dieticians in the dormitories would fall into this category. The other consists of a number of part-time and non-professional personnel service, such as the work done by faculty advisers, dormitory sponsors, housemothers, etc. It would seem to be advisable to let the person in the administrative position under whom these activities fall set up stand-

ards for them which are in accordance with the goals of the program of the particular institution.

We have also omitted from our classification the position of Director, Dean, or Coordinator of all personnel services. We would recommend that he be chosen from the ranks of persons trained in accordance with one of the special programs outlined above. Thus he would be sure to satisfy all the basic requirements in addition to possessing the particular administrative talents such a position requires.

If these recommendations were adopted by the Association there would yet remain the question of their effectuation. The committee did not consider recommendations along these lines to fall within its scope. It is the sense of this group, however, that certification of college personnel workers for these variously defined positions should not be contemplated at the present time (Related to this question is that of the status of psychologists within the ACPA who will be certified by the American Psychological Association Board of Examiners for educational personnel work.)

We do not feel that it would be helpful to the growth and integration of our profession for a set of standards, if adopted, to be used as a basis for rating personnel workers now actively engaged in college and university work. These standards are not equated for the value of years of experience in this field which have made some of the members recognized authorities who may lack the formal training specified in these recommendations.

It is hoped rather that as standards are set up they will become criteria for the training and selection of college personnel workers now in training or who will enter training in the future, and as a guide to personnel administrators for the selection of college personnel workers in our schools in the future.

MINORITY GROUPS IN OUR SCHOOLS

EDWARD J. SPARLING

President, Roosevelt College of Chicago

I SHOULD like to state at the beginning that the absence of quotas in the admission of minority groups into Roosevelt College has not created unusual headaches for our administrators. There is no basis whatsoever for hysterical fears about interracial marriage, spread of social diseases, lowered academic standards, or the spread of irresponsibility. We are a perfectly normal community of six thousand men and women, dedicated to respect people as people, without regard to the color of the skin or the tenets of their religious creed. Our student problems in general are of normal scope and intensity. Our standards of honesty, cleanliness, health, industry, and attitude indicate that minority groups do not bring with them a wave of unusual social, moral and physical problems. To contend that they do would be absolutely contrary to the facts as we have experienced them.

I do wish, however, to bring to your attention three specific problems which minority groups in our schools must face, and to indicate how we at Roosevelt College are dealing with these problems:

1. The problem of discrimination and undemocratic administrative control.
2. The problem of inequality of achievement.
3. The problem of restricted opportunity in job placement.

The Problem of Discrimination and Undemocratic Administrative Control

It is difficult to determine the exact influence which boards of control have upon the details of school administration because no systematic study has ever been made of exactly what goes on inside of board meetings. It is also difficult to prove the hypothesis that minority group problems in a school flow, in

part at least, from the kind of administrative control which the school has. A recent study did indicate, however, some interesting facts concerning the boards of control of our thirty-one leading universities:

1. The control of our universities is almost entirely in the hands of the top 9.5 per cent of our socio-economic group as classified by income.

2. That men in proprietary, legal, and managerial occupations represented 48 per cent of the boards of control in 1860, and that representation increased to 73.6 per cent of the boards of control in 1930, when the last survey was made.

3. When polled on the question of academic freedom, 34 per cent of the trustees acknowledged in writing that the scope and freedom of teaching should be restricted.¹

But, we are still left with this question: Exactly how does control of schools by the top socio-economic classes create problems for minority groups in our schools? I do not mean to suggest that race prejudice and high socio-economic status have positive and significant correlation. Certainly, we recognize that some of the bitterest and most irrational prejudices are on the lower socio-economic levels. I am suggesting, however, that the quota system is, to say the least, related to the type of upper middle-class mores which have dominated our schools since the nineteenth century. I am further suggesting that one of the greatest problems the Jewish, Negro, or other minority-group student faces, whether he is in high school or college, is the quota system if he wishes to continue his education. I am suggesting that this oppressive atmosphere of a quota system poisons and warps social, political, and moral attitudes and creates serious guidance problems among minority group students.

The quota system tries to be rational and humane by paying lip service to the American ideal that the underdogs (at least some of them) deserve a chance; it argues that quotas are democratic because student bodies must be "representative." Quotas seem plausible because they are expedient; they seem non-vicious because they are subtle and allow a token type of

¹ Beck, Hubert, T. *Men Who Control Our Universities*. New York: King's Crown Press, 1947. pp. 61-62, 112.

interracialism. At the bottom of the quota system lies the assumption that minority groups are inferior; without this assumption there would be no objection to a non-quota system based on merit. The quota system is a kind of "gentlemen's agreement" which is a strong characteristic of the mores of the upper middle-class citizen in America. It is the so-called "best citizens," as we measure them by socio-economic status, who control our schools and therefore impose the quota system on peoples they consider inferior.

If members of minority groups are fortunate enough to gain admission to schools of higher education, they face two problems closely related to the quota system:

First: Many schools and colleges have an administrative hierarchy which is a model of authoritarianism.

Second: Negroes and Jews face an elaborate set of social discriminatory practices on college campuses.

The President's Commission on Higher Education recognized the danger in an undemocratic administrative hierarchy when it stated:

To achieve practice in democratic action the President's Commission recommends a careful review of administrative policies in institutions of higher learning; revision may be necessary to give students every possible experience in democratic processes within the democratic community. Young people cannot be expected to develop a firm allegiance to the democratic faith they are taught in the classrooms if their campus life is carried on in an authoritarian atmosphere.²

In order to wipe out the problems of an authoritarian administrative hierarchy (which is a problem for all students) and the problem of social discrimination against minority groups on the campus, I suggest the following:

1. That faculty members be given a greater voice in administrative control.
2. That the philosophy and basic aims of the school be so stated that a complete air of freedom exists on the campus.

As a contribution toward this first goal, Roosevelt College sought to establish a new kind of board of trustees. A striking

² "Establishing the Goals." *Higher Education for American Democracy*. Vol. I. Report of the President's Commission on Higher Education. Washington: U. S. Government Printing Office, 1947, p. 14.

feature of our Board is the large voice exercised in it by the Faculty. Of the twenty-one members of the Board, five are elected by faculty members from the Faculty. Furthermore, the Board is composed of representatives of business, labor and the professions; the Board is also interracial and intercreedal. It truly represents all the main segments of life in the Chicago community. With this organization, no segment of the Board can control the academic and administrative processes in line with its particular ideas. So far as we know there are only two Negroes serving as trustees of any college which has white students. Both of these Negroes are on our Board. We know only four other colleges which give labor a voice on their boards of control. Of the some 30,000 men and women controlling our institutions of higher learning, we know of only seven persons from the ranks of labor. There is only a token representation of Jews, and women, the largest American minority group, are poorly represented, even on boards of coeducational institutions. Is it unreasonable to suggest, therefore, that the inadequate representation of large segments of the American people has been a cause of minority group problems in our schools?

As to the actual administration of Roosevelt College, the President and the Deans must request a vote of confidence from the Faculty at the end of each third year. The "yes" or "no" vote is by secret ballot and its purpose, in the case of the President, is to convey to the President and through him to the Board, for its guidance, the sentiment of the Faculty. The vote of confidence on the Deans has a similar purpose. Our Departmental Chairmen are elected for terms of three years by the President, the Dean of the particular school and members of the Executive Committee after the voting members of the department have filed their "recommendation." Members of the teaching staff are appointed by the President after hearing the recommendation of the Dean of the school in which the appointment is sought and of the Chairman of the Department and its voting members. The Faculty enjoys the right of recalling any officer it elects. Furthermore, any full-time employee of the College has the right to file a grievance and a formal grievance procedure is outlined by which complaints

eventually may be carried all the way to the Board of Trustees, if necessary.

We believe that this is a democratic administrative structure which gives us an unparalleled opportunity for free inquiry into any controversial area. Thus we feel that we have eliminated restriction, either direct or indirect, which might be exercised by a conservative board of control or by the vested interests of an entrenched administrative hierarchy.

As to the creation of a greater atmosphere of freedom, the only answer is to establish a school which refuses to count people or consider them on any other basis than as people. The recent stand taken by the National Student Association is commendable. It is high time that other discriminatory practices among students are abolished. Some of the most vicious discrimination is found in the fraternity system. Our students are now taking the lead in this matter and on April 1 will hold at Roosevelt College a conference of some twenty-six independent fraternities which are interested in a new national fraternity based upon democratic principles. Our Faculty Constitution, student government, and all student clubs operate on the basic principle of equality. The result is well reflected by a statement which was spontaneously made to a reporter of the *Louisville Courier-Journal* when he interviewed a member of our Faculty upon the occasion of the recent inauguration of President Taylor at the University of Louisville. Our faculty member made this statement:

At Roosevelt College there is a feeling of living in a free world which can be matched in few schools in this or any other country, as far as I know them. This does not mean that every faculty member and every student is liberal, tolerant, or a near-god. But there is no sense of fear. There is assurance that your colleagues who differ with you also respect you as an individual and as a fellow citizen. Students, too, differ with each other and with some of us from time to time. But freedom from fear is here, and freedom of speech is here, and freedom of assembly is here.³

The problem of discrimination is, of course, not exclusively the product of undemocratic administrative control in our schools. It arises out of the many complexities of the American

³ *Louisville Courier-Journal*, February 22, 1948.

social order. It is no longer questioned that a serious degree of racial and religious discrimination does exist in our schools. The evidence is now in. Most of us are sorry and ashamed at the picture which the evidence reveals. A large part of American educational practice stands convicted before the bar of democratic justice. The evidence shows that discriminatory practices are particularly vicious in the graduate schools of our colleges and universities. For example, in 1925, of all the graduates of the City College of New York who applied for admission to medical schools, 58.4 per cent were admitted. In 1943 this figure had fallen to 15 per cent.⁴

A recent study reveals that the greatest discrimination against minority groups occurs in the fields of dentistry, medicine, pharmacy, law, business administration, fine arts, architecture, veterinary medicine, and forestry. Minority groups are holding their own in the fields of social work, journalism, engineering, and home economics. Minority group representation has somewhat increased in the fields of osteopathy, music, theology, library science, education, military science, and nursing. The greatest discrimination at present seems to be in the professional law schools (where Jewish enrollment fell from 26% in 1935 to 10% in 1946).⁵

In 1945 when Roosevelt College was founded, two fundamental principles were stated in the Charter.

1. To provide a teaching faculty for such college which shall be both free and responsible in the discovery and dissemination of truth; and

2. To provide educational opportunities to persons of both sexes and of the various races on equal terms.

Today we have a great interracial student body of 5600. We do not know how many of what specific races we have represented because we do not count our students that way. This is a great step toward the achievement of a greater democracy in American higher education.

In many sections of the United States the problem of discrimination is primarily a political or legislative one. The

⁴ Kingdon, Frank. "Discrimination in Medical Colleges." *The American Mercury*, October, 1945, p. 4.

⁵ Baer, Max F. "Counting the Jewish College Students" *National Jewish Weekly*, November 11, 1947.

clearest example of this is, of course, the laws of most southern states which outlaw equal opportunity. The legality of this barrier has recently been tested in the Supreme Court of the United States. This was the well-publicized Oklahoma case. The Supreme Court ruled that states must provide equal educational opportunity for Negroes and whites. However, the now famous phrase "separate and equal" did not solve the basic problem because in a subsequent ruling the Court stated that separate and equal did not mean "identical." Thus, at the moment the matter of segregation of the races in southern schools is substantially where it was before the suit began. The results of such a system are also reflected in the amount of money which is spent in these states on white and Negro education. For example, in the District of Columbia three dollars are spent for white institutions for every one dollar spent for comparable Negro institutions, and in Kentucky this ratio rises alarmingly: In that State, \$42 are spent for white institutions for every \$1 spent for comparable Negro institutions.⁶

One avenue of attack on the problem of discrimination was the recent passage by the New York State Legislature of a bill outlawing racial and religious discrimination in the admission of students to private colleges, universities, and professional schools.⁷ The principle of legal coercion to end discrimination was opposed by the American Association of Colleges meeting in Cincinnati in January of this year. The reasoning of the college presidents who met in Cincinnati was that such legislation as was then being considered in New York would give the State an undue influence in the affairs of the schools. Instead, the American Association of Colleges felt that schools should voluntarily eliminate the undemocratic practices of discrimination in their admissions policies. Needless to say, I favor the stand taken by the State of New York, and I opposed the stand taken by the majority of the members of the American Association of Colleges. It seems to me that Roosevelt College three years ago and now the State of New York have taken the only possible stand for those who believe in the practice of

⁶ "Equalizing and Expanding Individual Opportunity." *Higher Education for American Democracy*. Vol. II, Report of the President's Commission on Higher Education. Washington: U. S. Government Printing Office, 1947, p. 31.

⁷ *New York Times*, March 14, 1948, p. 57.

democracy. This stand is to abolish all admissions quotas based on race and creed in determining who shall be educated in our schools.

The Problem of Inequality of Achievement

Commenting on the report of the President's Commission on Higher Education, the President of Fordham University recently stated:

By multiplying college facilities until they can care for every high school graduate who doesn't want to go to work, the Commission is not doing the colleges or the country any favor. How the Commission hopes to multiply the sheepskins and have fewer sheep, I cannot guess. This program threatens to suffocate us with tides of mediocrity.⁸

Although I do not agree with President Gannon, it would be less than honest to suggest that schools which aim at an education for the masses in large urban areas are free from worries about the "tides of mediocrity" which flow to the doorstep.

The special problems of minority groups in our schools are related to the general educational problem of discovering potential abilities and training them. Educational administrators in an institution such as ours must be especially alert to the possibility that poor socio-economic environment may have covered up potential abilities. We must also be aware that some deficiencies can be eliminated by special training while others cannot. Finally, we must also ask ourselves whether or not inequalities of achievement are due to any special characteristics of the various minority groups in our student body. Two complementary ways of dealing with the problem of unequal achievement have been initiated at our school. The first is the initiation of an Abilities Research Program, and secondly, the establishment of strong remedial programs in the fields of reading, writing and arithmetic.

The Abilities Research Program has been initiated with the help of Science Research Associates of Chicago. Lyle Spencer, the President of this organization is on the Board of Trustees of the College. Our Abilities Research Project has succeeded in

⁸Very Reverend Robert I. Gannon, President of Fordham University, as quoted in *Time Magazine*, February 23, 1948.

interesting some of the outstanding leaders in the field of testing and measurement of human behavior. Such authorities as Kuder, Thurstone, Warner and Kinsey are assisting us in the Abilities Research Program. We have taken a sample of some 400 students and have given them a battery of 36 different tests. Some of these tests, such as some designed by Thurstone, have not been given elsewhere and were especially designed for this program, as was a special device created by Warner for the establishment of socio-economic classification. Our immediate problem is to find out whether we can, through this comprehensive testing program, predict with any degree of certainty how well our students will do in college. Our student body represents a wider range of abilities than is generally found in a four-year college, and our large number of minority group students gives us an unparalleled opportunity to discover whether or not these minority groups have special remedial problems. We have just begun to correlate the results of the testing with the course grades obtained by the test subjects in their regular academic courses. Thus, we are not, at this time, able to report any significant conclusions. Our program to date has, however, indicated that we can profitably pursue further investigation based on the following hypotheses:

1. Potentiality (possibility of growth within the limits imposed by hereditary factors) is not distributed according to color lines, racial make-up, national boundaries or socio-economic class.
2. It is possible through research to discover those potentialities that are trainable and those that are relatively untrainable.
3. If an individual has a high proportion of both trainable and relatively untrainable ability-traits, then he is an excellent risk for college training regardless of previous achievement.
4. If an individual has a high proportion of untrainable ability-traits but a low proportion of trainable ability-traits, he is still a good college risk if a special effort is made through remedial programs to bridge the gap between achievement record and achievement potential. It is in this category that we find many of our minority group problems.

As to whether or not there are special problems among mi-

nority groups we are not yet in a position to say. Our own experience in the field of the interrelation of environment and intelligence is that there is no scientific evidence which would prove that minority group members, such as Negroes, are less able to profit by a college education than white students. This view is substantiated by Dr. George Stoddard who has declared:

Perhaps all that present day intelligence testing can do in connection with the racial problem has been done: It has revealed no basis in scholastic aptitude for discrimination against any race or nationality present in the American population.⁹

It has been stated that members of minority groups tend to lack social know-how and are often unfamiliar with the handling of highly verbalized and abstract material. Sometimes this has led to premature assumptions concerning the nature of intelligence. At other times this has caused many educators to group members of minorities into types and special groups. A recent study by the American Council on Education indicates that this typing and grouping has had an unfortunate repercussion in the writing of American's textbooks for the elementary and secondary level. This study came up with the following conclusions concerning the treatment of minority group problems in the text material of our schools:¹⁰

1. Intentional bias was not found in American teaching materials, but frequent value judgments and implications are often unconsciously or carelessly expressed; and even more pronounced, there are omissions of data and gaps in curriculum planning which result in failure of the teaching outlines and materials to come to grips with the issues especially connected with the problem of minority groups in our population.

2. The dignity and the worth of the individual are not stressed in most texts. The individual is usually submerged into a group; there is not adequate attention paid to the nature and value of human personality *per se*.

3. In treating group organization, too much emphasis is placed on "typical" group members such as Jew, Catholic,

⁹ Stoddard, George *The Meaning of Intelligence*. New York: McMillan Company, 1944, p. 247.

¹⁰ In *The Educational Record*, published by the American Council on Education in Washington, D. C., July, 1947, pp 235-262

Negro, etc., and too little on the variety of individuals within groups.

4. Regarding the treatment of Jews, most of the U. S. text materials deal with information concerning ancient Jews before 79 A.D., and treat these people as if they had never changed. Many inaccuracies were also discovered in the description of Jews as "race"; there was very little recognition of religious, economic and cultural variation among Jews.

5. In treating immigrants, too much emphasis was placed on "out-groups" and there was too much over-simplification in referring to "new" (meaning after 1880) immigrants as alarming "hordes" or "swarms." These latter groups of immigrants were usually referred to as "problems."

6. In treating Negroes, scientific data about race were conspicuous by their absence. Very often the treatment of Negroes was on a "folklore" level.

It is easy to see how such text materials, circulating in our public schools today, do harm in connection with the problem of minority groups. Thus, a vicious and undemocratic cycle is maintained: Individuals are viewed as types, treated in text material as types and are studied as types rather than individuals. At Roosevelt College we treat everyone as an individual and not as a type or member of any special race. This philosophy has enabled us to eliminate completely the tensions which exist in some interracial schools. In a recent survey, one of our graduates rather neatly summed up the results of our experience by declaring, "I shall always look back with justifiable pride upon this fraternal association of the races of mankind upon the common ground of learning."

The Problem of Restricted Opportunity in Job Placement

Any school which concerns itself with the training of large numbers of minority group members will sooner or later discover, as we have, that minority group members need more-than-average development of skills if they are to be placed in jobs which are commensurate with their abilities. Our job placement experience to date shows that we do great service to minority group students by urging them to attain more varieties of skills than they might normally be expected to need. We

have counseled with them to add such skills as accounting, statistics, psychometrics, and typing even when these subjects are not their major interests. Very often these extra skills have been the factors which tip the scale in favor of a minority group member applying for a position with a borderline firm. One practical result has been our success in introducing Negroes into higher grade positions in those firms already employing Negroes in menial jobs.

Chicago is a crucial area in the development of principles of fair employment practice, and we believe that Roosevelt College is in a unique position to contribute to the establishment of such principles. Although we receive and train students freely, without regard to race or creed, it is inevitable that we would find that the business and industrial world does not act so freely and fairly in accepting minority group members for job placement. In a survey of a couple of hundred firms in Chicago, we found that only 29 per cent would accept the best qualified person for the job regardless of race, color or creed. The problem is general but we feel that our position in the community is unique and our approach is unusually effective.

We reject the negative approach of trying to use coercion on prospective employers. Our approach is to talk things over with key business and industrial personnel leaders, try to discover and to overcome problems of well-meaning employers, and to counsel intelligently with minority group members regarding their own adjustment problems and the key importance of their own job performance when and if they enter a firm with a non-discriminatory policy. We believe this is a constructive approach. Our immediate problem is to enlarge our activities in this field so that results can be followed up systematically and valuable data thus obtained for contributing to the development of principles of fair and intelligent employment practice. We have followed the practice of guaranteeing our recommended candidates and have enjoyed some success in convincing business, especially in the wholesale and industrial field, that a well-trained member of a minority group is a better personnel investment than a less capable native white American.

Summary

I have pointed out what can be done in three specific areas to promote greater democracy in higher education. It is axiomatic that colleges cannot transmit that which they do not possess. They cannot train young people for democratic living unless Boards of Trustees, Administrators, Faculties and student bodies live and act democratically. Colleges and universities must take seriously their high mission of providing democratic leadership for the strengthening and perfecting of our great country to the end that "government of the people, by the people and for the people shall not perish from the earth."

•

WHO SHOULD ATTEND OUR COLLEGES

HILDA DAVIS

Dean of Women, Falladega College

TODAY we are considering the question of members of the so-called ethnic minority groups in our school populations. The word "ethnic" is used in the sense of "pertaining to groups of mankind distinguished by common customs and characteristics." To have said simply "minority groups" would have been to invite the possibility of too many interpretations, for there are many kinds of minorities if one uses the dictionary meaning for "minority," "the smaller in number of two aggregates."

Wallace Spegner in his book *One Nation* identifies as "the unaccepted" in American life, Filipinos, Japanese-Americans, Chinese, Mexicans (or Latin-Americans as they are called in some parts of the country), the American Indians, Negroes, Jews, Catholics.¹ In as much as another session is to be devoted to a consideration of intercultural relations, it seemed wise to omit from the present discussion the problems of minorities of various origins. In many schools the Catholics are definitely not in the minority, nor are they easily identified. Of Spegner's minorities we have left for consideration the American Indians, Negroes, and Jews. The American Indians, of whom there were 377,000 in the United States in 1943, are confined on various reservations and only occasionally are members of this group found in schools and colleges other than those operated especially for them. It appears, then, that Jews and Negroes are the groups which claim our attention particularly, since their youth comprise the largest number of ethnic representatives to be reckoned with in our educational systems.

The question of the acceptance of minority groups into the colleges of the nation is one fraught with high emotional tension,

¹ Spegner, Wallace. *One Nation*. Boston: Houghton Mifflin Company, 1945. Chapter 7, p. 141.

particularly in regard to the acceptance of Negroes. Consider the vitriolic editorials appearing in newspapers in certain states following the recent Supreme Court decision in the case of *Ada Fisher versus the State of Oklahoma* and the report of the President's Commission on Higher Education. The meeting of the governors of nine states to consider a regional university for Negroes, the various plans being devised and practised in several states to provide "separate but equal" educational facilities for Negroes, also are not so "mute evidence" of the concern of many people in the matter. In other states where the exclusion of Negroes from the schools is not accomplished by law it is effected by such means as quota systems or the excuse of capacity enrollments already accepted, no matter at what time of the year a Negro applies. Even in states with a Bill of Rights, such as Ohio and Illinois, one finds private schools whose announcements carry such a statement as "We regret that Under The Academy's rules, negro students cannot be accepted."²

The situation in regard to Jews is much more subtly handled. Quota systems operate without being announced. The result is that colleges and schools which do not use such methods find themselves almost segregated schools, so heavily Jewish is their enrollment; for example, City College of New York, about 90 per cent of whose students are Jewish.

I assume that the members of the Program Committee in asking me to participate in this session knew my attitude on the subject. It is my firm conviction that our enrollments in both public and private colleges should comprise representatives of all groups; that students should be chosen on the basis of their native endowment, their academic preparation, their ability to profit by the type of education available at a particular institution, and not on the basis of their race, color, creed, or family background. I believe that having accepted students on this basis an institution should grant to all students the opportunity of participating freely in any of the activities provided and of using all of the facilities without restrictions based on race. To educate our students to function adequately as citizens of our society our schools must make it possible

² The Chicago Academy of Fine Arts. *Bulletin*, 1947, p. 27.

for the youth of the nation to know each other in such a way as can come only through constant cooperation and participation in activities of real and mutual interest.

I am fully aware that my convictions are far from being realized in various parts of the country, not just in the South and Southwest where state laws have operated to keep separate Negroes and Caucasians, and sentiment to exclude Jews from the schools. At the same time, my further remarks have pertinence, I believe, even for those persons who at present have no members of ethnic minorities in their colleges. The problems facing members of ethnic minority groups are much the same as those of any other minorities in an inclusive student population.

In as much as the personnel officers of a college are usually the group who come most directly in contact with all of the students and frequently have the deciding voice in many matters involving members of minority groups, first of all college administrators, deans, counselors, or whatever the title, must examine their attitudes; they must look squarely at their own stereotypes and prejudices. Do we believe that all Negroes steal; that Negroes are untidy and dirty; that they are immoral; that they are mean and treacherous; that they lack the capacity to do high-level academic work; that most Negroes have syphilis; that all Negroes can sing; that Negroes are childish, light-hearted, and perpetually cheerful; that Negroes, as the saying goes, "would rather be with their own kind?" How many of us have accepted the prevailing stereotypes regarding Jews? Do we think of Jews as clannish, unfair and addicted to cheating especially in money matters, as so aggressive that to advance their own interests they ruthlessly override others, as unclean and unkept in their persons and premises? Undoubtedly some of these accusations are true about some Jews and some Negroes, but so are they true about some Caucasians. The reverse is also true of some Jews and some Negroes. There are many members of these ethnic minorities who are meticulously clean, who are governed by high ethical and moral standards; many who have demonstrated their capacity for intellectual achievement on a level far beyond the understanding of the average one of us; many who are affable, loyal,

generous and thoughtful of others. Are we guilty of generalizing from too few cases? It is unfortunate that we have made for ourselves models for Jews and Negroes. We have taken characteristics of individuals and have constructed from these traits types which we call *the* Negro and *the* Jew. Now, as we face the problem of further integration of minorities into the various aspects of American life and more immediately into the American colleges, we must rid ourselves of the fallacious tendency to type people and to react to them as groups. Rather, we must seek to know and to evaluate them as individuals. William Lloyd Ines in writing the "Role of The Christian College" made the statement,

It is quite the logic of things to begin by hating a man for one thing and to end up by finding many other things which make us dislike both that man and others, too. Hatreds never run singly; they go in devilish and deadly packs. And racism is easily pulled into the pack of human enemies which are created by an evil and selfishly competitive order.³

The role of the administrator and the counselor is an important one. Frequently the final decision as to whether there shall be members of minority groups among the students in a college or how far they may participate in college activities rests with the President, the Dean, the Dean of Women, the Resident Head. During the twenty years in which I have worked in the Deep South time after time I have shared with both Negro and white students the frustration they have experienced when a dean of women or a college president forbade the participation of white students in some interracial activity.

That the decision was that of one individual is demonstrated by a case in point. The daughter of one of my friends was a member of the Home Economics Department of a publicly supported university in a state which has a Bill of Rights. As this student was an out-of-state student, she was not familiar with the conditions prevailing in the Practice House of the Home Economics Department. Unwittingly then, she registered for this required part of the curriculum. Later she discovered that this requirement was modified for Negro

³ Nelson, William Stuart, editor. *The Christian Way in Race Relations*. New York: Harper and Brothers, 1948, p. 133.

students. Imagine her astonishment and chagrin to find herself the only student living in one side of a double house while the teacher and some nine or ten white students lived in the other side. I was told that for the required number of weeks this student, living alone, went through the motions of cooperative living. She would be invited to dinner in the other side of the house occasionally; then the head of the Practice House would help the Negro girl to prepare a dinner to which the white students were invited. When this student's family protested, the President said in effect that he would not require any white students to live with Negroes. The only point in citing this case is to indicate that as individuals we frequently make decisions on the basis of our prejudices and impute these decisions to student desires.

No longer can we evade the issue with the statement, "The students are not ready for this." Early in September, 1947, about 800 students from colleges throughout the United States, among whom were many representatives of southern colleges, organized and formally adopted a constitution from which I quote a portion of the preamble,

We, the students of the United States of America, desiring to maintain academic freedom and student rights, to stimulate and improve democratic student governments, to develop better educational standards, facilities, and teaching methods, to improve student cultural, social, and physical welfare, to promote international understanding and fellowship, to guarantee to all people, because of their inherent dignity as individuals, equal rights and possibilities for primary, secondary, and higher education regardless of sex, race, religion, political belief or economic circumstances . . . do hereby establish this constitution to the United States National Student Association.⁴

Another large group of American students assembled as the United Student Christian Council adopted the following Resolution on Racism, a part of which I quote,

We have been confronted, at this Fourth Annual U.S.C.C. meeting, with a broken world desperately groping for a way of life which can assure peace and hope. . . .

Racism is a cancer that eats deep into our democracy, the Christian Church and even our own student fellowship. It

⁴ *Radcliffe Re-News*. Cambridge, Mass., February, 1948, p. 2

cuts across culture, isolates individuals and dims the Christian witness. In this festering sore in the very heart of what we try to do as Christians, the Negro in our life is hurt, frustrated and wounded and the souls of all citizens are blighted by this disease of our social order. Worst of all, the witness of the Christian Church is weakened because the racism breaks its fellowship and thus the Church itself makes explicit the negation of Christian brotherhood. The problem of racism within the Church in our nation is not sectional. It is a corporate guilt of all the Church in all parts of the nation.

Among the resolutions accepted with the above statement was

We call upon each member of U.S.C.C. to urge his national student Christian movement within the coming year to make one of its major program emphases that of improving race relations.¹

To come even nearer the problem--ten years ago the University of Missouri students voted in large numbers for the admission of a Negro, Lloyd Gaines, to that institution. More recently a majority of the students at the University of Oklahoma voted for the admission of Ada Fisher to the University of Oklahoma law school. In Auburn, Alabama, the students were polled on their attitude toward the admission of Negro students to the state-supported polytechnic institution located in that city. Twenty-eight per cent of the students expressed themselves in favor of admitting Negroes and eleven per cent as indifferent. While the larger number voting in this instance expressed a negative attitude, we must recognize that students are becoming concerned about the problem and are developing attitudes against the exclusion of members of minority groups.

I can not emphasize too much the importance of an objective attitude toward members of minority groups held by deans and counselors, and to that end I would urge that they avail themselves of opportunities to know members of the various ethnic groups individually. Courses and institutes concerned with human relations and more specifically with intercultural relations such as are being offered at Mills College, the University of Chicago, Harvard and Radcliffe Colleges, Yale University—to name just a few of them—would be valuable to deans and

¹ *Federation News Sheet*. Monthly Bulletin of the World's Student Christian Federation, February, 1948, p. 5.

counselors in changing their points of view as they are faced with the problem of minority groups in the colleges. In a course in Behavior Problems several years ago Dr. Mandel Sherman frequently reiterated his belief that people do not change their attitudes until they have had different experiences from those which produced the attitudes which need changing. Before we try to guide immature students into constructive attitudes and behavior in regard to minority groups, we, ourselves must have wholesome attitudes and courage to face the various problems involved.

The problems to be dealt with by deans and counselors concerned with ethnic minorities are those which are usually identified with the profession—vocational and educational counseling, housing, social and other extra-class activities, personal problems, and inter-personal relations generally, although these problems may be intensified when race is a factor. Educational and vocational counseling can scarcely be separated, for practically all students in colleges are looking toward some vocation after undergraduate- or graduate-school days. After the general educational level they choose courses which will move them toward their vocational goals. Too many times Jewish and Negro students even in high schools are steered into courses which fit the various counselors' stereotypes of vocational opportunities for a particular ethnic group, rather than into courses which satisfy these students' interests and abilities. For example, at the college at which I have worked for the past eleven years, year after year we have enrolled graduates of a high school in a certain large northern city. All of these students have been graduates of technical curricula rather than of the general liberal arts course. When we inquired into this circumstance because a number of these students of high intellectual ability showed evidences of personal maladjustment and faulty attitudes toward school work, we learned that the counselors in this high school, all of whom were white, invariably enrolled Negro students in the technical courses because they believed that the girls could find work only in home economics activities and the boys in mechanical trades. One of our graduates living in this city had believed that all of his seven children could not be identical

in abilities and vocational interests. He sent them and a number of the children of his friends to his alma mater for us to try to correct the results of the faulty counseling which these young people had received in high school. It is true that there are still a few vocational activities to which Negroes and Jews are denied access and some others in which the participation of these ethnic groups is severely limited. In counseling such students the counselor should of course point out the limitations, but he should not be too negative in his attitude toward any vocation toward which a student aspires, provided the student has ability and the other factors which seem to predict success in that field. Who knows but that the very student now being counseled will be the first of his group to perform in a certain field? Negroes are now faced with numerous vocational opportunities heretofore closed to them and find that there is none of their group prepared to hold a certain position, not because no Negro has the ability to do that work successfully but because no Negro has prepared to do it. The National Urban League and the Jewish Occupational Council have published materials which are useful in vocational counseling with members of these ethnic groups.

What of housing? Is it better to house members of a particular minority group together, i.e., in a separate building, in one wing or on one floor? Should roommates always be of the same ethnic group? My answer to both of these questions is "No." Very definitely, members of minority groups should not be separated from the other students. Such segregation arouses resentment and inferiority feelings in the minority group students. It also tends to increase the suspicion and lack of understanding of one group toward the other. Certainly there are some white students who do not want to associate with Negroes and/or Jews. The housing practices of a college should be made according to the aims and ideals of the college. Practically all colleges claim to be educating students for democratic living. The needs of the individual white students who are not sufficiently mature socially to live next door or across the hall from Negroes or Jews, for example, should be accepted by the dean or the

counselor and adjustments made in the individual cases. It should be possible to provide all white neighbors and room-mates for those who can not yet accept persons of a different ethnic group.

Experience has shown that frequently persons who have thought their prejudices immovable have been thrown by unavoidable circumstances into contacts with members of ethnic minorities and have found the experience not repulsive, as they had anticipated, but interesting and later even pleasant. During the years of my residence in International House and in several of the dormitories of the University of Chicago I have frequently been told by some of my associates who had not identified me racially of their anticipated reactions at the prospect of living with Negroes, eating at the same tables, using the same bathrooms and the like, and of their surprise that the actual contact had not affected them in the way they had anticipated.

In this situation, as in most situations, the Counselor, the Resident Head, the Personnel Dean set the pace. One head resident I know used to invite one or two Negro girls to eat at her table each meal during the first days of the quarter and introduced them to other students. She made it possible for girls who had negative attitudes toward Negroes to discuss them frankly and without censure. The same rapport existed between her and the Negro girls. She recognized what many others do not, that the battle is only half won when white students change their attitudes toward minority groups. Sometimes an even more difficult part of the problem is to convince the members of ethnic minorities that the overtures of white persons are genuine expressions of respect, appreciation, and acceptance. Too frequently Negroes, especially, have been so deeply hurt by the frustrations and indignities which they have experienced at the hands of others because of race that they withdraw into themselves and actually repel the efforts of others to become acquainted with them.

Members of ethnic minorities should not be excluded from activities as a matter of policy. At the same time, they should not be chosen for positions simply to have a group represented, even though the individuals available are not interested in the

activity or lack qualifications for the particular positions. A stratagem of persons who wish to demonstrate their prejudice regarding a Negro's lack of intelligence is to put him in a position for which he is not fitted. His subsequent failure is then generalized to other members of his group.

Of course, all students should be able to use all of the facilities of the institution, including the swimming pool. Several years ago one of my fellow deans of women told me that the Negro students in her college were not permitted to use the swimming pool which the other students used by arrangement with a YWCA in her mid-western city because of the high rate of venereal disease among Negroes in that city. I have always wondered why the health service in her college was unable to determine which Negro students might be free of venereal disease.

What kind of campus jobs will be given students of the various ethnic groups? Should Negro students be appointed only to janitorial or domestic jobs or will they be given work opportunities commensurate with their skills and abilities? This aspect of a student's functioning in a college is almost entirely in the hands of administrative officers.

Another problem which must be faced in considering the inclusion of minority groups in a student enrollment is the extent to which they shall participate socially. If social activities are a valid part of the education of students, and we have long ago agreed that they are, then all students must be free to participate in them. This area represents the one about which many people have the greatest anxiety. Will not interracial fellowship lead to fraternizing and love affairs? Marriage across racial lines is the "red herring" which is dragged into all discussions of activities inclusive of minority groups. It is indeed the tool of the narrow-minded and the fearful. This is a rather groundless fear. In colleges where members of ethnic minority groups attend and participate freely and normally in the activities there have been few if any cases of interracial marriage. Efforts to keep white youth from association with Negroes, for example, has heightened for both Negroes and white youth the sense of glamor and air of adventure in mutual association. Last year during my

residence at the Chicago International House the social relationships of Negro and white students were frequently discussed. Several southern white women students who for the moment were "going with" Negroes, to use the student vernacular, said more than once, "I've never known Negroes before, but now I find that they are not at all like the stories I've been told. They're just like anybody else." When the novelty of the situation had worn off, these white students could accept Negroes as "just like anybody else" and not confine their attention to them. Guidance by a wise counselor can help young people to realistically wholesome social relationships with members of other ethnic groups.

Interpersonal relations are involved in all aspects of a student's life in a college. If he must help to support himself, he is in contact with a supervisor and with other workers. In the dormitory obviously there are many persons with whom he must adjust. The responsibilities of counselors as well as other personnel workers may be greatest in this regard. If a student under emotional stress can talk to a counselor he trusts to listen without censure or praise to the expression of private and hostile attitudes, the student has made at least one step toward ridding himself of his negative attitudes. The therapeutic process will be advanced if the counselor can clarify for the student the feelings being expressed and at the same time refrain from telling him how he should react or what the next steps are. No one but the student himself can really decide what is best for *him*, although superior knowledge and experience may give a counselor insights which the student has not yet achieved.

Opportunities for group discussion and face-to-face expression of attitudes and concern will be of great service to members of ethnic minorities as well as to Caucasian students in our colleges.

If I had taken a text for my remarks today, it would have been, "Physician, heal thyself." I believe that the serious problems feared by those who hesitate now to have an inclusive student group will not materialize if those who are responsible for the guidance of the students—and I include in this category faculty and staff from the president on—are

committed to the policy of accepting persons as individuals and of providing opportunities for them in which they may develop according to their capacities and interests. No matter how many students are on a campus and how busy counselors and deans are, their attitudes toward minority groups soon become known. Let each of us, then, examine our attitudes and do something constructive toward developing an understanding of an appreciation for all people. Then we can disapprove of what some individuals do, like some members of a group and dislike others, understand why particular persons act as they do. When we have achieved this level of emotional objectivity ourselves, we can help our students to function more effectively in our society which is unquestionably a mixed one of 377,000 Indians,⁶ 4,500,000 Jews,⁷ and nearly 14,000,000 Negroes to mention only groups in the United States not identified specifically with another nation.

⁶ Spegner, *op. cit.*, p. 141.

⁷ *Ibid.*, p. 301.

PUTTING OCCUPATIONAL INFORMATION ACROSS¹

ARTHUR H. BRAYFIELD

Dean of Student Personnel, Long Beach City College

IN 1909, Frank Parsons devoted forty of the one hundred and sixty-five pages of his counseling classic, *Choosing A Vocation*, to the topic of occupational information under the general heading of "The Industrial Investigation." Since that time practically all professional personnel workers have stressed the importance of such information in the counseling of both youth and adults. All of the standard texts in the field refer to the use of occupational information as an integral part of any student personnel program. Several authors have written extensive treatments of the topic, the most recent of which is Shartle's book titled *Occupational Information* which I have elsewhere referred to as heralding "the coming of age of what Kitson has tentatively labeled for want of a better term the 'Science of Occupationology'."

Yet, despite this early recognition of the importance of occupational information in counseling and related personnel activities and its acceptance as one of the major tools of personnel workers, one searches the literature almost in vain for any systematic consideration, on other than a techniques level, of the use of occupational information. Descriptions of specific media and materials abound, but discussions of basic principles underlying their application are conspicuous by their absence.

Before we try to "put occupational information across" I believe that it is essential that we critically examine the customary uses of occupational information and attempt to formulate a few relevant principles to guide any discussion of

¹ Discussants whose comments are not recorded are Dwight A. Stewart, Personnel Supervisor, Radio Corporation of America, Indianapolis; Barbara H. Wright, Senior Consultant in Counseling, Public Schools, Minneapolis; Max F. Baer, National Director, Vocational Service Bureau, B'nai B'rith, Washington, D. C.; Margery E. Hoppin, Research Assistant, Western Personnel Institute, Pasadena

specific techniques. The remainder of my remarks are devoted to such an end.

At the outset I should like to make two important distinctions. The first is one made by Grace Munson and Lester Schloerb writing in the recent *A Basic Text for Guidance Workers*, edited by Clifford Erickson, wherein they differentiate between the *extensive* and the *intensive* use of personnel materials. For example, the extensive use of occupational information materials means that "throughout all the years of the school life of the child, there should be an effort to emphasize the real nature of the world of work." However, they also point out that there are certain elements which demand concentrated study and activity often at a particular point in the student's career, and this they term the intensive use of such materials. In this paper I have considered only the intensive use of occupational information.

A second necessary distinction is between the use of occupational information in group instructional situations and in individual counseling. (Parenthetically the use of the word "individual" is redundant when associated with the term "counseling" but I shall use it for emphasis.) Both aspects are discussed in this paper.

Group instruction in occupational information as a part of a school program ordinarily is organized in one of three ways: (1) as a pervasive part of the "regular" curriculum; (2) as a special unit in standard school subjects such as English, Civics, etc.; or (3) as a course organized specifically for the purpose. My concern is with the latter form of organization only.

The instructional methods used in such courses may include such diverse techniques as lecture-discussion, textbook and other printed material study, outside speakers, motion pictures, classroom dramatizations, radio, special projects, careers booklets, and case studies.

I should like to contrast two group instruction approaches now current. The first is what I prefer to call the *standard occupational information approach* (the term "standard" carries a less evaluative connotation than the term "traditional" which might well be substituted). This approach is associated with the names of such personnel workers as Brewer, Billings,

Forrester, and others. Its characteristics can be treated under the headings: point of view, subject matter, methods, assumptions, and evaluation.

The basic point of view of this approach appears to be that the acquisition of information about the world of work is the proper starting place for vocational planning.

The subject matter typically is concentrated upon occupational information *per se* with lesser emphasis upon psychological materials. The following quotation from Billings in her *Group Methods of Studying Occupations* is a clue to the relative emphasis. "There are a number of occupational considerations," says Billings, "which should be treated *briefly* (*italics by the present writer*). Students should know that there are certain basic skills such as mechanical ability, clerical and commercial skill, verbal ability and others." The Brewer and Landy text, *Occupations Today*, is perhaps representative.

As to methodology, much of the work is classroom and textbook centered. Workbooks which stress the accumulation of occupational information are frequently used. A common device is the Careers booklet. With respect to the latter, I cannot refrain from commenting that my experience with college freshmen has led me to formulate the concept of "vocational planning arrested in its development at the level of the ninth-grade Careers booklet" by which I mean that many entering freshmen have fixated upon vocational plans arrived at as a result of the apparently traumatic experience of writing up some vocational field in order to satisfy the requirements of a ninth-grade course in Occupations. Occasionally such plans have some slight degree of appropriateness for the individual concerned.

Although I have found in the literature no formal rationale for the standard occupational information approach, I am inclined to believe that the following assumptions are at least implicit.

1. The individual who has been formally exposed to occupational information will make more appropriate vocational choices than the individual who has not been so exposed.
2. The individual will make specific application of the con-

tent of the course to his own situation with a minimum of individual assistance

- 3 The individual completing such a course will develop attitudes and skills which will carry over as new problems of vocational adjustment arise.

None of these assumptions has been tested. The evaluation of this standard approach has been largely in terms of the amounts of information which have been gained as a result of exposure. The typical experimental design is to test an experimental and a control group at the beginning of a course in Occupational Information with a test which stresses occupational information. At the end of the period of instruction both groups are re-tested. The differential, of course, is the exposure of the members of the experimental group of the course. The comparison, with the proper control for initial status, is then made between the status of the two groups on re-testing. The results have been conflicting. A survey of studies by such investigators as Billings, Jessup, Hand, and Williamson leads to the conclusion that exposure *may* lead to the accumulation of more occupational information. The more adequately designed studies do concur, however, in the finding that if you set out to teach occupational information, you will produce gains in the amount of such information possessed by the members of the instructed group. I leave to you the interpretation of the significance of this finding.

The second major group instruction approach is what I prefer to call the *socio-psychological* or what Munson and Schloerb of the Chicago Public Schools term the *self-appraisal and careers* type of course. This approach is associated with the work of Neuberg at Wittenberg College, Allen and Towne at Providence, the Chicago program, the work of the Minnesota group including McLean, Williamson, Paterson and Hahn, and others including Margaret Bennett at Pasadena and the more recent development at Philadelphia.

The basic point of view of this approach is that the acquisition of information about the self is the proper starting place for vocational planning.

The subject matter of this approach strikes a balance between the psychological study of the individual and the study of

occupational information. In contrast with Billings' statement which recommended only brief attention to the study of individual characteristics, Munson and Schloerb state that "Every effort must be made, however, to show a well-planned distribution between self-appraisal study and career study." Aptitude and interest patterns provide the scheme for the presentation of occupational information. The printed materials from the Chicago program are perhaps representative.

As to methodology, the emphasis is upon using both the classroom and the community as important learning laboratories. Projects involving considerable student initiative and activity are basic to a successful program. Rather than the standard Career booklet, the intensive case study becomes one of the important vehicles for translating the findings of the work of the course into meaningful materials to be applied to a given individual. The case study technique leads directly into the individual counseling which is considered the *sine qua non* of the whole approach.

The following assumptions underlying this approach are interpretations primarily of the writings of the Providence and Chicago people and are based also on my own association with the program at Minnesota.

1. The individual who has been formally exposed to the work of such a course will make more appropriate vocational choices than the individual who has not been so exposed.
2. The number of individuals making appropriate vocational choices will increase as individual counseling is provided to supplement the group instruction.
3. The individual completing such a course will develop attitudes and skills which will carry over as new problems of vocational adjustment arise.

To my knowledge, there has been only one attempt to rigorously evaluate the socio-psychological approach. This is the study by Harold Stone of the vocational orientation program in the General College of the University of Minnesota. Stone found, as had some of the evaluators of the standard occupational information approach, that students exposed formally to such information made greater gains in measured

committed to the policy of accepting persons as individuals and of providing opportunities for them in which they may develop according to their capacities and interests. No matter how many students are on a campus and how busy counselors and deans are, their attitudes toward minority groups soon become known. Let each of us, then, examine our attitudes and do something constructive toward developing an understanding of an appreciation for all people. Then we can disapprove of what some individuals do, like some members of a group and dislike others, understand why particular persons act as they do. When we have achieved this level of emotional objectivity ourselves, we can help our students to function more effectively in our society which is unquestionably a mixed one of 377,000 Indians,⁶ 4,500,000 Jews,⁷ and nearly 14,000,000 Negroes to mention only groups in the United States not identified specifically with another nation.

⁶ Spegner, *op. cit.*, p. 141.

⁷ *Ibid.*, p. 101.

PUTTING OCCUPATIONAL INFORMATION ACROSS¹

ARTHUR H. BRAYFIELD

Dean of Student Personnel, Long Beach City College

IN 1909, Frank Parsons devoted forty of the one hundred and sixty-five pages of his counseling classic, *Choosing A Vocation*, to the topic of occupational information under the general heading of "The Industrial Investigation." Since that time practically all professional personnel workers have stressed the importance of such information in the counseling of both youth and adults. All of the standard texts in the field refer to the use of occupational information as an integral part of any student personnel program. Several authors have written extensive treatments of the topic, the most recent of which is Shartle's book titled *Occupational Information* which I have elsewhere referred to as heralding "the coming of age of what Kitson has tentatively labeled for want of a better term the 'Science of Occupationology'."

Yet, despite this early recognition of the importance of occupational information in counseling and related personnel activities and its acceptance as one of the major tools of personnel workers, one searches the literature almost in vain for any systematic consideration, on other than a techniques level, of the use of occupational information. Descriptions of specific media and materials abound, but discussions of basic principles underlying their application are conspicuous by their absence.

Before we try to "put occupational information across" I believe that it is essential that we critically examine the customary uses of occupational information and attempt to formulate a few relevant principles to guide any discussion of

¹ Discussants whose comments are not recorded are Dwight A. Stewart, Personnel Supervisor, Radio Corporation of America, Indianapolis; Barbara H. Wright, Senior Consultant in Counseling, Public Schools, Minneapolis; Max F. Baer, National Director, Vocational Service Bureau, B'nai B'rith, Washington, D. C.; Margery E. Hoppin, Research Assistant, Western Personnel Institute, Pasadena.

specific techniques. The remainder of my remarks are devoted to such an end.

At the outset I should like to make two important distinctions. The first is one made by Grace Munson and Lester Schloerb writing in the recent *A Basic Text for Guidance Workers*, edited by Clifford Erickson, wherein they differentiate between the *extensive* and the *intensive* use of personnel materials. For example, the extensive use of occupational information materials means that "throughout all the years of the school life of the child, there should be an effort to emphasize the real nature of the world of work." However, they also point out that there are certain elements which demand concentrated study and activity often at a particular point in the student's career, and this they term the intensive use of such materials. In this paper I have considered only the intensive use of occupational information.

A second necessary distinction is between the use of occupational information in group instructional situations and in individual counseling. (Parenthetically the use of the word "individual" is redundant when associated with the term "counseling" but I shall use it for emphasis.) Both aspects are discussed in this paper.

Group instruction in occupational information as a part of a school program ordinarily is organized in one of three ways: (1) as a pervasive part of the "regular" curriculum; (2) as a special unit in standard school subjects such as English, Civics, etc.; or (3) as a course organized specifically for the purpose. My concern is with the latter form of organization only.

The instructional methods used in such courses may include such diverse techniques as lecture-discussion, textbook and other printed material study, outside speakers, motion pictures, classroom dramatizations, radio, special projects, careers booklets, and case studies.

I should like to contrast two group instruction approaches now current. The first is what I prefer to call the *standard occupational information approach* (the term "standard" carries a less evaluative connotation than the term "traditional" which might well be substituted). This approach is associated with the names of such personnel workers as Brewer, Billings,

Forrester, and others. Its characteristics can be treated under the headings: point of view, subject matter, methods, assumptions, and evaluation.

The basic point of view of this approach appears to be that the acquisition of information about the world of work is the proper starting place for vocational planning.

The subject matter typically is concentrated upon occupational information *per se* with lesser emphasis upon psychological materials. The following quotation from Billings in her *Group Methods of Studying Occupations* is a clue to the relative emphasis. "There are a number of occupational considerations," says Billings, "which should be treated *briefly* (*italics by the present writer*). Students should know that there are certain basic skills such as mechanical ability, clerical and commercial skill, verbal ability and others" The Brewer and Landy text, *Occupations Today*, is perhaps representative.

As to methodology, much of the work is classroom and textbook centered. Workbooks which stress the accumulation of occupational information are frequently used. A common device is the Careers booklet. With respect to the latter, I cannot refrain from commenting that my experience with college freshmen has led me to formulate the concept of "vocational planning arrested in its development at the level of the ninth-grade Careers booklet" by which I mean that many entering freshmen have fixated upon vocational plans arrived at as a result of the apparently traumatic experience of writing up some vocational field in order to satisfy the requirements of a ninth-grade course in Occupations. Occasionally such plans have some slight degree of appropriateness for the individual concerned.

Although I have found in the literature no formal rationale for the standard occupational information approach, I am inclined to believe that the following assumptions are at least implicit.

1. The individual who has been formally exposed to occupational information will make more appropriate vocational choices than the individual who has not been so exposed.
2. The individual will make specific application of the con-

tent of the course to his own situation with a minimum of individual assistance.

3. The individual completing such a course will develop attitudes and skills which will carry over as new problems of vocational adjustment arise.

None of these assumptions has been tested. The evaluation of this standard approach has been largely in terms of the amounts of information which have been gained as a result of exposure. The typical experimental design is to test an experimental and a control group at the beginning of a course in Occupational Information with a test which stresses occupational information. At the end of the period of instruction both groups are re-tested. The differential, of course, is the exposure of the members of the experimental group of the course. The comparison, with the proper control for initial status, is then made between the status of the two groups on re-testing. The results have been conflicting. A survey of studies by such investigators as Billings, Jessup, Hand, and Williamson leads to the conclusion that exposure *may* lead to the accumulation of more occupational information. The more adequately designed studies do concur, however, in the finding that if you set out to teach occupational information, you will produce gains in the amount of such information possessed by the members of the instructed group. I leave to you the interpretation of the significance of this finding.

The second major group instruction approach is what I prefer to call the *socio-psychological* or what Munson and Schloerb of the Chicago Public Schools term the *self-appraisal and careers* type of course. This approach is associated with the work of Neuberg at Wittenberg College, Allen and Towne at Providence, the Chicago program, the work of the Minnesota group including McLean, Williamson, Paterson and Hahn, and others including Margaret Bennett at Pasadena and the more recent development at Philadelphia.

The basic point of view of this approach is that the acquisition of information about the self is the proper starting place for vocational planning.

The subject matter of this approach strikes a balance between the psychological study of the individual and the study of

occupational information. In contrast with Billings' statement which recommended only brief attention to the study of individual characteristics, Munson and Schloerb state that "Every effort must be made, however, to show a well-planned distribution between self-appraisal study and career study." Aptitude and interest patterns provide the scheme for the presentation of occupational information. The printed materials from the Chicago program are perhaps representative.

As to methodology, the emphasis is upon using both the classroom and the community as important learning laboratories. Projects involving considerable student initiative and activity are basic to a successful program. Rather than the standard Career booklet, the intensive case study becomes one of the important vehicles for translating the findings of the work of the course into meaningful materials to be applied to a given individual. The case study technique leads directly into the individual counseling which is considered the *sine qua non* of the whole approach.

The following assumptions underlying this approach are interpretations primarily of the writings of the Providence and Chicago people and are based also on my own association with the program at Minnesota.

1. The individual who has been formally exposed to the work of such a course will make more appropriate vocational choices than the individual who has not been so exposed.
2. The number of individuals making appropriate vocational choices will increase as individual counseling is provided to supplement the group instruction
3. The individual completing such a course will develop attitudes and skills which will carry over as new problems of vocational adjustment arise.

To my knowledge, there has been only one attempt to rigorously evaluate the socio-psychological approach. This is the study by Harold Stone of the vocational orientation program in the General College of the University of Minnesota. Stone found, as had some of the evaluators of the standard occupational information approach, that students exposed formally to such information made greater gains in measured

knowledge than did a control group. He found also that work in the course produced a shift in vocational aspirations bringing them much more in line with reasonable expectation than was true for his control group.

Stone's most important finding, however, had to do with the effect of enrollment in the course upon the appropriateness of educational-vocational goals. Using this criterion of effectiveness, he discovered that both the control and the experimental group had lower proportions of optimal choices and higher proportions of poor choices at the end of the study. Thus the effect of the course appeared to be more disrupting than helpful. This seems to negate the first assumption.

However, and this is an extremely significant finding for personnel workers, when he compared a group of counseled students who had had the vocational orientation course with a group of counseled students who had not had the course, he found that the former group made significantly greater changes in the direction of appropriate choices than did the counseled group which had not had the vocational orientation background. I interpret this to mean that the values of this approach are more likely to be realized *when the group instructional program is accompanied by individual counseling* and, conversely, that counseling is likely to be most effective when group instruction has laid the necessary foundation.

It should be obvious that we lack sufficiently final evaluative data for either approach to be able to make definitive statements of basic principles. However, to express a personal point of view, I would postulate the following as guides to the use of occupational information *per se* in group instructional programs:

1. The emphasis should be distributed relatively evenly between psychological and occupational materials.
2. Study of the self should precede study of occupations.
3. Specific occupational information should be organized around aptitude and interest patterns rather than around traditional classifications such as the Census and the *Dictionary of Occupational Titles*.
4. Except for a *brief* overview of the entire world of work the occupational fields selected for group study should be

chosen with due respect not only to the expressed interests of the members of the group but particularly with reference to the types of occupations likely to be followed by the majority of the members of the group.

5. Each member of the group should review his work in the course with a trained counselor.

In essence, then, I am urging strongly that we abandon the standard approach to the use of occupational information in group instructional programs. I urge that we concentrate instead upon the development of more adequate forms of organization of group instruction which will integrate socio-psychological materials. Further, let us insure that individual counseling is a concomitant part of our group instructional program.

Turning now to the use of occupational information in individual counseling I find the literature sadly lacking in any systematic treatment of the topic. Rather characteristic is the mention, in a recent text, of "disseminating" occupational information to individuals. This lack of adequate (or any) rationale is perhaps symptomatic of the same type of thinking that clings to the standard occupational information approach in group instruction—that is, the failure to recognize the importance of individual diagnosis as pre-requisite to the application of occupational information.

As I understand it, modern scientific personnel work requires that we attempt to specify, as Williamson has stated recently, "Under what conditions and with what types of problems and individuals will a particular counseling technique facilitate the desired adjustment." To date I have found only one such attempt to specify the circumstances under which occupational information is used other than to "disseminate" it. Ruth Strang, writing in the *School Review* in November, 1945, illustrated the differential use of occupational information according to the position which a counselor occupies along a nondirective-directive continuum. This article represents an interesting attempt to clarify the role of occupational information in the present straw-man controversy with respect to non-directive-directive counseling.

A few years ago in organizing a new course in Occupational

Counseling, I attempted to formulate some differential applications of occupational information; I am presenting them here not as finished products but as illustrative of the possibilities for specifying the conditions under which occupational information may be used.

I have distinguished, for emphasis, between the *informational*, the *re-adjustive*, and the *motivational* applications of occupational information in counseling.

Perhaps the most frequent and the best understood use of occupational information is, as the name implies, simply *to inform*. Consider the circumstance where the counselee has arrived at an appropriate and satisfying vocational choice based on an accurate understanding of his own potentialities but is lacking detailed information about the field of his choice. The role of the counselor simply is to furnish the detailed information which is lacking.

Another distinctive circumstance is one which calls for confirmation by the counselor. This occurs when the counselee again has made an appropriate choice on the basis of adequate personal and occupational information but seeks to resolve his lingering indecision through a confirmation interview with the counselor. The counselor simply supplements the information or reinforces the client with the knowledge that his information is accurate. I am assuming, of course, that this is a straightforward occupational information case and that the counselor is reasonably certain that this is not an instance where psychological "weaning" is indicated.

A still different circumstance obtains when the counselee seeks to resolve his indecision resulting from a conflict between what appear to be two equally appropriate goals. Here the counselor's role is to supply highly specific information regarding each alternative as one possible basis for the counselee to make a difficult choice.

Finally, in some instances where immediate job placement is the primary consideration, the choice of occupation must be made in large part upon the highly specific information which the counselor furnishes.

The second classification is the *re-adjustive*. In my own experience in both school and community counseling centers

this has been one of the most frequently identified conditions for the differential application of occupational information. The major circumstance is the use of occupational information with the counselee who has set markedly inappropriate goals for himself either in terms of level of aspiration or field of work. In terms of technique, the counselor first uses leading questions regarding the nature of the occupation or field which the counselee has chosen. In turn, the counselor provides accurate information which may enable the client to gain insight into the illusory nature of his thinking when he finds that his conception of the occupation or field does not fit the objective facts. At this point the counselor usually is able to turn the interview to a consideration of the realistic bases upon which sound occupational choices are founded. It then becomes possible to re-orient the counseling situation to make a re-evaluation by the counselee of his personal characteristics the initial phase of his vocational planning. The essence of this counseling use of occupational information is the attempt to help the counselee evaluate the reality basis of his planning. This application is, of course, a more dynamic one than the strictly informational and probably calls for more skill on the part of the counselor.

Other counseling uses of occupational information are *motivational*. The first set of circumstances which I have identified I mention with some hesitation in Chicago. As you know, Carl Rogers postulates that "counseling can be of help only when there is a certain amount of psychological distress." Some of us, however, have forced upon us by parents and/or administrators the task of "provoking" our clientele to serious thinking about the problem of vocational choice. That is, it becomes incumbent upon us to gain the client's recognition and acceptance of the fact that eventually it will be necessary to face the problem of making a living, in consequence, the problem of vocational choice. Actually, then, we must seek to create the "psychological distress" about which Rogers writes. Under these circumstances (with which, apparently, Rogerian counselors do not need to concern themselves) I suggest that counselors use those types of occupational information such as the number of different ways of making a living,

the increasing difficulty which young people face in getting established vocationally, the complexity of the world of work, etc., which may help to arouse an interest in vocational planning.

Frequently the client conceives the problem of vocational planning to consist primarily of the acquisition of information about occupations without reference to personal and individual qualifications. In this situation the counselor furnishes sufficient information to "hold" the client's interest until the client perceives that he must have information about himself as well as the world of work. Thus the counselor provides the information demanded while he works to re-orient the counselee's thinking about the nature of the problem of vocational adjustment.

This "holding" usage applies also to the counselee who wants to be told what to do. The counselor carries him along by providing some occupational information while he works to help the counselee become more independent.

A somewhat different use of occupational information is to "arouse" the client's interest in using a so-called hidden talent. That is, occupational information is provided in order to build up identifications in those instances where, through testing and a re-evaluation of personal history data, the client has discovered unknown talents or interests appropriate to fields which he has not previously considered.

A related usage involves the "maintenance" of motivation. This occurs particularly in school situations where the curriculum engaged in often seems unrelated to later vocational activities. The counselor uses occupational information to strengthen the counselee's identification with the members of a given occupational group. One of the most useful, specific techniques is to get the counselee started in reading the trade or professional journals in the field. For an example close to home, I have encouraged graduate students in personnel work to affiliate with the American College Personnel Association and the National Vocational Guidance Association in order to build up their identification with the field. However, this situation can get out of hand unless skillfully handled. Increasing familiarity with actual occupational requirements

may increase the counselee's dissatisfaction with his educational diet.

Undoubtedly there are other ways of ordering and classifying the differential counseling applications of occupational information. My concern is that we cease to disseminate occupational information. Rather we should seek to specify with increasing exactitude under what conditions and with what types of problems and individuals we will use a particular type of occupational information.

Essentially, all that I have said involves one fundamental principle which is basic to any use of occupational information whether group or individual. Stated succinctly that principle is:

Any use of occupational information should be preceded by individual diagnosis.

UNESCO AND EDUCATIONAL RECONSTRUCTION¹

HAROLD I. SNYDER

Director, Commission for International Educational Reconstruction

"What can we, the pupils of Plunkett High School, do to promote world understanding?" "How can our Camp Fire Girls group help European children?" "Can our women's club help in any way in promoting the objectives of UNESCO?" "What can a group of university professors at Mainline University do to help our colleagues in China?" "Is there anything that we (a group of youth organizations) can do toward peace?"

The above questions represent samples of the thirty to forty requests which pour daily into the offices of the Commission for International Educational Reconstruction (CIFR). This is America speaking, the true heart of America at its humanitarian best. These people have no axe to grind, no political motivations, no selfish intent. They are genuinely concerned for their fellowmen, for the plight of youth in distant lands who must share with our youth responsibility for building a peaceful world. They want to share directly, somehow, in building a peaceful world, in promoting international understanding and world cooperation.

Fortunately, there are many ways in which they can share directly in the healing of the scars of war. One of the most important of these ways is by undertaking projects for international educational reconstruction projects which will not only help to re-establish educational opportunity for the youth

¹ Following this presentation, Harry D. Kitson, Professor of Education, Teachers College, Columbia University, lead a panel discussion on the Vocational Guidance Needs in Other Countries. Discussants were: Robert Carey, Director of Guidance, Yonkers, New York; Jean P. Jordann, Assistant Housemaster, South African College High School, Capetown; Mohamed Abdel Salam Ahmed, Director of Vocational Guidance Department, Ministry of Education, Cairo, Egypt; Witold Kruk-Olpinski, Adjunct Professor, University of Warsaw, Poland; Havard Skirbekk, Instructor, High School, Hurva, Norway; Edward G. Williamson, Dean of Students, University of Minnesota, and students from local universities.

of war-torn lands, but which will also strengthen UNESCO.² The initials UNESCO stand for United Nations Educational, Scientific and Cultural Organization, the most ambitious undertaking to date looking toward world-wide cultural cooperation. It is necessarily a complicated organization with 40 member nations, an international staff of 500, headed by the noted British scientist, Julian Huxley, headquarters in the old Hotel Majestic in Paris, and an array of projects ranging from the development of world literacy, the holding of seminars for international understanding, the establishment of science research centers, the translation of the literature of the resistance movement during the war, the establishment of international book exchanges, and others too numerous to mention. Foremost among all UNESCO projects, fundamental to all UNESCO activities, is educational reconstruction. For without the restoration of educational opportunity, and of cultural institutions in the countries ravaged by war the ability of Europe and Asia and part of Africa to participate in long-term projects of international collaboration is strictly limited.

The speaker has heard this point made eloquently and repeatedly at five international conferences during the past two years. He has heard educators, scientists, scholars of the devastated countries describe their struggles against fearful odds to restore the cultural life of their nations. They stress particularly their problem of reclaiming and educating for useful citizenship the so-called "lost generation" of Europe and Asia. These are the millions of youth who during the war years were either deprived of schooling or were subjected to the perverted education imposed by conquering armies, youth still bearing the physical and psychological scars of war, youth who had been taught as a matter of patriotic duty to lie, cheat, steal and even to kill.

Upon these youth, who have known from early childhood the horrors of war and political turmoil, whose birthright has been hunger, cold, and personal insecurity, the devastated countries must depend for the leadership of the future. They will share

² Several specific projects were outlined by the present speaker in his remarks on Council Day at the 1947 convention in Columbus. These suggestions, which appeared in the May issue of *Occupations*, are still valid and will not be repeated.

with our youth responsibility for building a peaceful world. If denied an opportunity for education, for development of those skills and attitudes essential to permit them to play their part in the reconstruction of their countries, if disillusioned and embittered by a postwar world which seems to have no place for them, they may again become prey to totalitarianism. Deprived of opportunity, they constitute a potential threat to the security of our more fortunate American youth. They are, therefore, of concern to us both because we Americans are humanitarians, concerned with helping suffering humanity everywhere, and because we see that helping them is enlightened self-interest.

UNESCO challenges us to face this problem of educational and cultural reconstruction. It is merely one aspect of the wider problem of facing the essential, obvious fact that the world is no longer divisible into distinct geographical and political areas. The implication of this basic fact has not yet sufficiently impressed itself upon Americans—even upon American educators. Phrases such as the "air-age," the "atomic age," "world interdependence" are too often dismissed as mere slogans. How seldom do we face realistically the human implication of these phrases, the irrefutable fact that what happens to human beings in any part of the world is of direct concern to those in all parts. Widespread insecurity, frustration, desperation—particularly of youth—in the devastated countries is definitely our concern.

In November at Mexico City the second Conference of UNESCO reaffirmed the high priority given a year earlier to educational reconstruction. Full and generous recognition was given by Dr. Huxley and by the delegates of the various countries to the splendid cooperation which UNESCO has been receiving from American organizations—appreciation not only for the size of the contributions now totalling approximately 150 million dollars but also for the spirit of unselfish and sympathetic understanding which has characterized most of them. It was clearly brought out, however, that while the response of voluntary organizations has exceeded expectations, the need is still appalling. Continuing economic difficulties and political unrest have so delayed general and educational reconstruction that another year of intensive effort is essential.

Fortunately, as compared with the first UNESCO Conference in Paris, fewer countries reported in Mexico City an utter lack of simple school supplies such as pencils and notebooks. These are still needed in Eastern and Southern Europe and in the Orient. There is, however, a serious general lack of educational and scientific equipment, of books and the means of producing them, and of trained personnel. There is as great a need as before for combining moral and spiritual aid with the provision of material assistance.

To fill in the gaps which are necessarily left in what is essentially a voluntary effort, the conference at Mexico City authorized a very limited program of direct operations by UNESCO. A few items of technical equipment of certain types are being purchased from military surpluses. Recognizing that trained personnel is the basis of sound reconstruction, UNESCO's exchange of persons program will stress the needs of the devastated countries. A special study of the problems of war-handicapped children is to be made. A book coupon plan to permit the purchase of foreign books in domestic currencies is being studied. Voluntary service projects and youth camps are being encouraged. A limited fund has been set aside to send two or three educational advisory missions to the devastated countries.

The conference also called upon the United Nations, its specialized agencies and all member states to see to it that all programs of economic reconstruction include appropriate emphasis upon training and education. The CIER has already carried out this mandate by urging the President and Congress to include such emphases in the Marshall Plan and in other governmentally financed reconstruction activities.

Reconstruction is clearly the one all-UNESCO concern which permeates the work of every section and every project of the organization. The implication is clear that at the present time the best way and almost the only way in which organizations can play a direct and immediate part in the promotion of UNESCO's long-term objectives is to undertake practical projects in response to the Organization's reconstruction appeal.

International cooperation is too often something which we merely talk about. The real test of our sincerity lies in what

we do. Our colleagues abroad judge us by our actions. Reconstruction activity is one method of action which succeeds not only in closing the gap between verbalization and action, but also leads to direct and satisfying contact between organizations, institutions and individuals in this country and similar groups abroad. Through such projects literally millions of American school children, teachers, church members, club members, community leaders and professional workers of all kinds have helped to make a direct contribution toward international understanding and world peace. Following the various suggestions in the CIER's publications, they are sending books and other materials, offering fellowships and study grants, developing advisory missions, participating in voluntary service projects. This is a functional approach to international understanding.

The year 1948 may well go down as one of the most critical years of decision in world history. Those of use who believe in education must keep constantly before us the fact that—essential as they are—the provision of basic necessities, of economic assistance and of military assistance, the dependence upon bilateral governmental efforts and upon diplomacy are not alone sufficient to bring about the recovery and to secure the peace of the world. Believers in education do not need to be told that war has its beginnings in the minds and spirits of men. These aspects of human existence must be nourished as well as the physical aspects. Our organizations and our institutions are demonstrating through their reconstruction projects a concern for the minds and hearts of youth still suffering from war-imposed handicaps. They are demonstrating through concrete voluntary action that Americans have a concern for the individual as well as for the mass. They are keeping before them the double objective of meeting urgent needs abroad and strengthening UNESCO and the other United Nations agencies upon which we must depend to provide the machinery for lasting international cooperation.

COUNSELING STUDENTS FROM OVERSEAS¹

CLARENCE LINTON

Adviser to Students from other Lands, Teachers College, Columbia University

Introduction

THERE are at present approximately 400 counselors of students from overseas in our colleges and universities. These counselors are generally called Foreign Student Advisers. They are responsible for counseling a large number of students from almost every part of the world. In addition to the counselors in higher institutions, hundreds of persons in binational foundations and agencies, in professional and religious agencies, in embassies and consulates, here and abroad, are concerned with the purposes and processes of promoting intercultural understanding through the exchange of students.

It is probable that the counselors of students from overseas in the U. S. A. will organize a national association at their meeting to be held in Ann Arbor in May, and, in due course, I hope this association may be affiliated with ACPA.

But our discussion today needs a broader context. At the outset, I choose to give it the broadest possible context. I assume that we are all desirous of achieving *One World*, and that we believe that exchange of students should contribute something toward the achievement of this ultimate goal. I fear that we are not sufficiently aware, however, of the opportunity we have of establishing *One World in Our Front Yard*.

We in the U. S. A. have an unprecedented opportunity and, therefore, an unprecedented responsibility. Much of our world looks to us for leadership today, fearful that we will betray their trust, yet hopeful that we shall justify their faith. The colleges and universities of the war-devastated countries were largely destroyed. In no other land are higher institutions prepared to receive large numbers of students from over-

¹ Discussants whose comments are not recorded are Forrest G. Moore, Adviser to Foreign Students, University of Minnesota; Robert M. Strozier, Dean of Students, University of Chicago.

seas. Students from most other parts of the world are seeking ways and means of coming to the U. S. A. It is estimated that 70,000 students from overseas are now waiting for such an opportunity.

There is a long standing and widespread belief that exchange of students is one important means of achieving *One World*. Comenius, the Moravian bishop and educator, wrote in 1643:

There is needed in this country an immediate remedy for the frenzy which has seized many men and is driving them in their madness to their mutual destruction. For we witness throughout the world disastrous and destructive flames of discord and wars devastating kingdoms and peoples with such persistence that all men seem to have conspired for their mutual ruin which will end only with the destruction of themselves and the universe. Nothing is, therefore, more necessary for the stability of the world, if it is not to perish completely, than some universal rededication of minds. Universal harmony and peace must be secured for the whole human race. By peace and harmony, however, I mean not that external peace between rulers and peoples among themselves, but an internal peace of mind inspired by a system of ideas and feelings. If this can be attained, the human race has a possession of great promise.²

Comenius would have promoted these lofty ideals by means of a Pansophic College whose scholars from all lands might gather to arrange the elements of knowledge needed for mental understanding among mankind.³

Many others have dreamed of and planned for intellectual cooperation and intercultural understanding. Many individuals have acted on this faith. Between 1810 and 1910 about 10,000 Americans studied in German universities. Some 50,000 Chinese students have come to the U. S. A. during the past 75 years. Before the outbreak of the World War II the University of Paris had a total annual enrollment of 40,000 students, of whom 10,000 were from other lands. The American Association of University Women, the various professional societies, the religious bodies, the International Rotary, a number of bi-national foundations and agencies, and many other private groups are currently engaged in the promotion of student exchange. Raymond B. Fosdick, President of

² Kandel, I. L. *Intellectual Cooperation*. New York: Teachers College, Columbia University, pp. 77-78.

³ Carr, William G. *Only by Understanding*. New York: Foreign Political Association, 1945, p. 30.

Rockefeller Foundation, in his Review of the work of the Foundation for 1945, said:

In the last thirty years, both directly and through representative national agencies in various countries the Rockefeller Foundation has given fellowships to approximately 7,700 men and women and has spent for this purpose more than \$20,000,000. These 7,700 fellows have come from 72 different countries and have represented many races, creeds, backgrounds and branches of scholarship. Although scattered throughout the world, they have shared a common experience and they speak the common language of humanism and science (p. 43)

Since World War I most of the major nations have instituted intercultural programs. France and Germany have spent as high as one-fifth of their foreign office appropriations for this purpose. Britain and the U. S. A are late comers in this field. The British Council was established in 1934. Our Division of Cultural Relations was established in the Department of State in 1938; its duties are now performed by the Division of International Exchange of Persons and the Division of Libraries and Institutes in the Office of International Information and Educational Exchange.

"Education for international understanding" is one of the four major projects of the United Nations Educational, Scientific and Cultural Organization to be accomplished, in part, through a drive to stimulate more fellowships, scholarships, and travel grants for the exchange of teachers, students, and research workers. UNESCO has appropriated funds for a program of reconstruction fellowships during the current year which will provide a period of study and travel for trained persons from China, Czechoslovakia, Denmark, Greece, the Netherlands, Norway, the Philippines and Poland. . . . It is estimated that some 30 of the 48 fellows will come to the United States.⁴

Our government has recently taken positive action toward a program of exchange of persons. All ACPA personnel should be informed regarding this program which is based on three acts of Congress.

I. *The Fulbright Act*, passed last year, was an amendment to the surplus property act. Some of the important pro-

⁴ Source: Letter from Mr. Laurence Duggan, Director, Institute of International Education, 2 West 45th Street, New York 19, New York.

visions of and decisions taken under this act are important for our discussion.

- A. Separate agreements are now being negotiated between each participating country and the United States. Agreements have already been made with China and Burma, Australia, Austria, Belgium, Czechoslovakia, Egypt, Finland, France, Greece, Hungary, Iran, Italy, the Netherlands, Netherlands Indies, New Zealand, Norway, the Philippines, Siam, Turkey, India, Pakistan, and the United Kingdom are eligible.
- B. Countries participating must set aside in their own currencies an amount agreed upon, not exceeding \$20,000,000, to be spent at the rate of not more than \$1,000,000 per year over a 20-year period, for exchange of persons.
- C. Funds are to be used to pay the expenses of citizens of the United States in the participating countries and the travel expenses of nationals of the participating countries to and from the United States, provided such expenses are payable in the currencies of the participating countries.
- D. The President's Foreign Scholarship Board has made a broad interpretation of "exchange of persons" to include: professors, teachers, students, and specialists in arts and sciences.
- E. The Institute of International Education has been designated to coordinate the exchange of elementary and secondary school teachers. The Conference Board of the Associated Research Councils has been designated to coordinate the exchange of professors, lecturers, research workers, technicians, and other specialists. The U. S. Office of Education will administer programs for sending teachers to public schools under the Fulbright Act and the American Council on Education will do the same for private schools.
- F. Colleges and universities, professional groups, and other private agencies will be invited to participate in regional screening committees.
- G. Foundations will be established in the participating

countries to coordinate services to citizens of the United States and to supervise the screening of nationals of participating countries seeking to come to the United States.

- H. It is estimated that the total funds available under this Act may be \$140,000,000, which amount, supplemented by provisions of the G. I. Bill of Rights, may conceivably provide opportunity for 5,000 citizens of the U. S. A. to study abroad each year for the next twenty years.

II. *The Smith-Mundt Act*, approved by the President, January 27, 1948, authorizes the Secretary of State:

- A. "To provide interchanges on a reciprocal basis between the United States and other countries of students, trainees, teachers, guest instructors, professors, and leaders of specialized knowledge and skill."

- B. To carry on a genuine educational program with the peoples of the world parallel with but separate from the *Voice of America*.

- C. To provide for exchange of persons, information, films, books, pamphlets in education, arts and sciences.

The Smith-Mundt Act has been passed and approved, but appropriations have not yet been made for its exchange provisions. Whereas the Fulbright Act is limited to designated countries, to currencies of those countries, to twenty years, and mainly to citizens of the United States; the Smith-Mundt Act will apply as broadly to the world as circumstances permit, will provide United States dollars, will apply to nationals of the participating countries as well as to citizens of the United States, and "can operate as long and as richly as Congress wants it to."⁵

It is hoped that the two acts will supplement each other. "For example: a Burmese student might get his transportation to the United States paid under the Fulbright Act (with Burmese rupees). Once in the United States,

⁵ Source. Horace Mann League Letter Number 43, 1201 Sixteenth Street, N W, Washington 6, D C, February 23, 1948.

his living and tuition expenses might be paid under the Smith Mundt Act (with United States dollars)."⁶

III. *The G. I. Bill of Rights*

The timeliness of this discussion is further attested by the fact that the G. I. Bill of Rights grants governmental subsidies to ex service personnel for study in most of the higher educational institutions of the world. It is assumed that programs made possible by these three Acts will be coordinated.

This is merely a sketch of the larger context in which I wish to discuss the counseling of students from overseas. There is another consideration of greater immediate importance. It is perhaps naive to assume that exchange *per se* promotes intercultural understanding. Those of us who have been abroad with our military forces know that experience abroad may intensify prejudice rather than promote understanding. We know, too, that some students from overseas go home with no genuine understanding of our culture. Something more than exchange of bodies is required. There must be careful selection of the *unofficial ambassadors* who are exchanged. They must be prepared in mind and heart for such study. And they must be dedicated to a mission of service in their own countries to give purpose and meaning to their experiences abroad. Much understanding must be put into the policies, plans, and processes of exchange.

Here is where the counselors of students from overseas come in. It is their special responsibility to study the problems and to learn the processes and techniques by which all of us may become more effective in our relationships with these unofficial ambassadors.

This will require, among other things, a reorientation of our personnel officers, our faculty members, and our students. The job cannot be done by doing things *to* and *for* students from overseas. We must learn to do things *with* these students. We must receive them as *informants*, as *consultants*, and as *unofficial ambassadors*. Knowledge of their families, their educational systems, their way of life, their problems, their hopes, and their aspirations will greatly enrich our understand-

⁶ *Ibid.*

ing. There are many ways in which doing things *with* these students will enrich our lives as counselors, as instructors, as fellow students, as friends.

This is the most effective way to assist students from overseas in achieving their purposes in coming here. Once we demonstrate our genuine interest in them and their countries, they become interested in seeking from us information about our families, our schools, our way of life, our hopes, and our aspirations.

I judge the success of our program of counseling at Teachers College, Columbia University, in large part by what I observe taking place spontaneously in the cafeteria lines, in the library, in the corridors, on our visits to schools and other institutions. Are students from overseas teaming-up with our students in clubs, in study groups, in theater parties, on hikes, on shopping trips, in the social rooms, in the dormitories, and in vacation periods in American homes?

Some important considerations which I shall not have time to develop are. (1) Experience of the counselor of student from overseas confirms the critical importance of making available to all students an integrated personnel program, (2) a well-selected group of students from other countries should enrich the educational environment for all students, (3) a clear bid should be made for a greater sharing of responsibility with the community.

I. Present Status of Exchange Students in the U. S. A.¹

- A. There are approximately 20,000 students from overseas in the U. S. A.
- B. These students come from 110 different countries.
- C. There are approximately three men for each woman
- D. Fully 85-90 per cent of these students come on family funds, or are working part-time to meet expenses. Approximately 10-15 per cent of them hold government scholarships, or scholarships from private agencies. Not more than one-tenth of them are receiving some

¹ Source: *Unofficial Ambassadors*. New York. Committee on Friendly Relations Among Foreign Students (347 Madison Ave.), 1947-1948

financial aid from our colleges and universities. Many are having a very difficult time financially.

F These 20,000 students seek admission and tend to enroll in larger and better known universities.

F The largest numbers of students come from the following countries:

1. Canada	3,522
2. China	2,622
3. India and Pakistan	1,214
4. Mexico	704
5. Cuba	672
6. Philippines	622
7. Norway	552
8. Turkey	521
9. France	470
10. Greece	385
11. Palestine	376
12. Colombia	374
13. Brazil	354
14. Panama	320
15. Iran	312
16. Peru	296
17. Egypt	283
18. Netherlands	271
19. Venezuela	268
20. Poland	256

G. They are largely concentrated in the larger universities as indicated by universities which enrolled 100 or more in 1946-47:

1. Columbia University
2. University of California
3. Harvard University
4. University of Michigan
5. New York University
6. University of Minnesota
7. Massachusetts Institute of Technology
8. Cornell University
9. Louisiana State University
10. University of Washington

11. University of Wisconsin
12. Howard University
13. University of Chicago
14. Syracuse University
15. University of Texas
16. Yale University
17. University of Illinois
18. University of Pennsylvania
19. Purdue University
20. Ohio State University
21. Iowa State University
22. University of Detroit
23. Fordham University

H. These 20,000 students from overseas are distributed over the length and breadth of the U. S. A. attending more than 900 colleges and universities.

I. They are chiefly interested in our technology: engineering, science and agriculture.

J. There are approximately 400 Counselors of Students from Overseas, some of whom are present in this group today. In a real sense all members of ACPA should be members of this group, because all college personnel officers have a responsibility for this program, as do our faculty members and students.

K. Each of these students from overseas has problems and needs the assistance of a counselor. Most of them have innate resourcefulness and manage fairly well, but some need very special assistance with personal problems.

1. A Hindu student arrived from Bombay by air after an elapsed time of 48 hours in clothing appropriate for a semi-tropical climate. At the time of his arrival New York was experiencing semi-Artic climate. He had no living accommodations, little money, and was late for registration.

2. Truth is sometimes stranger than fiction, as is the story of one of our Chinese women, who saw two of her sisters murdered by a Japanese soldier, who killed the soldier and escaped into the interior, came to the U. S. A. in 1946, developed t. b., and is now in a sanatorium where she must stay for at least eighteen months.

3. A student from the Gold Coast, West Africa, after long

planning and preparation finally managed to come to the U. S. A. For a full year after his arrival he was unable to obtain any money from his sponsor. After his money finally came through he discovered that it would take three years more than he had planned to complete the requirements for the doctorate.

II. Functions of the Counselor of Students from Overseas

It is important that we view the functions of the counselor in the context of the total program of information, orientation, and counseling. These functions are interdependent. It is the responsibility of the counselor to establish a total program of exchange of students in his institution, which is so conceived, so planned, and so coordinated that each student will obtain the assistance he needs at the time he needs it. Much of this program will be accomplished by others, but the counselor must take cognizance of what others can and should do, he should utilize available resources and supplement them as needed.

A. The first need of the prospective student from overseas is for information necessary for intelligent choices. Intercultural understanding requires a much better job in matching interests, abilities, and needs of individuals, countries, and institutions. More specifically, the information needed by the prospective student includes:

1. Information about the countries offering the best opportunities for study in his chosen field. This is the first requirement. Should he go to England, France, India, China, Brazil, or the U. S. A.? This is mainly a problem for the country from which the prospective student comes. But the counselor of students from overseas must understand it and contribute what he can to its solution. He must prepare himself for counseling American students who plan to study abroad, because he is a natural source of information. By way of example, consider for a moment where our ex-service personnel are going on the G.I. Bill of Rights. They are going, or trying to go, mainly to countries and institutions best known in the U. S. A.: Oxford, Cambridge, Paris. They are almost completely neglecting excellent opportunities, in India, China, and South America.

Our information literature sent to inquiring persons abroad should point up this problem and suggest sources of first-hand information, such as: fellow countrymen who have traveled and studied in different countries, cultural institutes or foundations where they exist (and many more may be established in the future), the cultural attaches of embassies, and consuls of the different countries.

2. Information about the institutions best qualified to serve the student's needs in the country finally selected is the second requirement. There is too much concentration in a few institutions at present, both in the U. S. A. and elsewhere. It should be the general policy of governments and educational authorities, and therefore of counselors of students from overseas, to encourage and to facilitate a wider distribution of students, both to countries and to institutions within countries. In the long view this is wise policy for all concerned. The information literature we send to inquiring persons should point up the problem and make suggestions for its solution. Fellow countrymen who have traveled and studied in the country chosen, cultural institutes, cultural attaches, selection committees, and consuls should be helpful sources of supplementary information.

3. Information about how to prepare for the period of study abroad is the third requirement. The information obtained by students is largely hit or miss at present. We should greatly improve our flow of authentic information to prospective students from overseas immediately. One of our best sources of help in this matter is our present group of students from overseas. Most of them return to positions of leadership. Prospective students should be urged to confer with a countryman who has recently studied in the U. S. A., and preferably with one who has studied at the institution where he wishes to study. There is need for common use of a guide to students from overseas which will provide general information applicable to all. Such guides have been prepared by the Institute of International Education including a general orientation handbook entitled *Meet the U. S. A.* These should be used by all of us, until more comprehensive guides are available.

There is a second aspect of information of concern to the

counselor, namely, the information needed by his institution about the prospective student, his preparation, and his country. This is chiefly the responsibility of the Admissions Officer, but the counselor is responsible for seeing that adequate policies and procedures are followed in obtaining such information. He must know the sources of information and to whom to go for assistance. Every problem of admissions, whether quality of the student admitted, delay in sending the necessary certificate of admission, determination of equivalents, or requirements for degrees, will be brought to the counselor sooner or later. He should anticipate such problems. The unofficial ambassador should be carefully selected. No service is done the individual or his country, and almost certainly the major objective of intercultural understanding will not be accomplished, if unqualified students are admitted, or if the student is admitted to the wrong institution. Having said this, I hasten to say just as emphatically that the process of selection and admission must be conceived as a personnel service: We are dealing with persons—persons who are the only hope for accomplishing our objectives. Administrative convenience and efficiency are important, but they must always be subordinated to human values. I wish it were possible to require every admissions officer to experience just once what it feels like to be a stranger in a strange land. He would certainly be more understanding of the human factors involved in admissions. The counselor must make this his concern; otherwise, he will be overwhelmed with problems, many of which are quite unnecessary.

B. The second need of students from overseas is an orientation to American life and the plan of study in the U. S. A., especially in the institution in which he is to study. Here I ask your indulgence for a brief outline of the policies, procedures, and program of Teachers College, Columbia University. We have had a large number of students from overseas for nearly half a century. For a quarter of a century a special program of information, orientation, and counseling has been in operation.

1. The chief purposes of our program are:

a) To induct the student from overseas into the life

of the College as a self-sufficient, equal member of the student body, as quickly as possible. To this end all the administrative, counseling, and instructional services of the College are available to him on equal terms with students from the U. S. A. In addition, special assistance is given as needed throughout his period of study, chiefly to enable him to make optimum use of the regular services

b) To establish a relationship with students from overseas which facilitates their contribution to the enrichment of the life and work of the College community in the promotion of intercultural understanding.

2. The current orientation program includes:

a) An Orientation Week is held immediately preceding the registration period each session and new students are required to attend. The principal activities of this week are.

(1) Counseling.

(2) Lectures on New York City, Columbia University, Teachers College and the Plan of Study and Advisement.

(3) Visits to points of interest on the campus and in New York City, with a view of familiarizing students with their physical environment and the transportation system.

(4) Diagnostic English tests to determine the competence of students in the use of the English language and to provide a basis for placement in English courses and for limiting programs, if necessary.

(5) Social events (luncheons and a tea) to promote acquaintance among students. Faculty members, older students from overseas, and American students participate in these social affairs.

b) An orientation course in *American Culture and Education* is required of all students from overseas during the first session of study. This course includes:

(1) One half-day a week spent in observation of schools and other institutions.

(2) Extensive use of films on American Culture and Education.

(3) Readings, lectures, discussions.

(4) A luncheon each week, to which members of the faculty and selected students are invited as guest speakers.

c) An International Film Series is a weekly feature of the orientation program in which films of different countries are shown. Students from these countries serve as resource persons and leaders of discussion. These film showings are open to all students of the College. They provide one means of doing things with students from overseas.

d) An International Club, membership in which is open to all students of the College, is joint sponsor with a faculty committee of the International Film Series, of social programs, and of various other means of inducting students from overseas into the life of the College.

C. The third need of students from overseas is counseling. All students need counseling; some more, some less.

1. Students from overseas have special needs for counseling because.

a) They are strangers in a strange land.

b) They are unfamiliar with American life and education.

c) They often are limited in their use of English.

d) They often have special personal problems of finding suitable living accommodations, meeting the inflated costs of living, understanding the regulations of the Immigration and Naturalization Service regarding their visas, and knowing to whom to turn for assistance.

e) They should have an interest in understanding American life. To this end they need to make friends, share experiences and do things *with* students from the U. S. A. While such experiences should be as spontaneous and natural as possible, students from overseas usually need assistance in making the initial contacts. They often need the assistance of a counselor in matters of etiquette.

f) They need special counseling regarding academic matters: choice of field of specialization, choice of major

field adviser, choice of courses, use of the library, methods of instruction, requirements for degrees, standards, marks and regulations.

2. Policies and practice at Teachers College will serve as examples of one way to provide the counseling needed:

a) When possible, information is sent by mail to the prospective student well in advance of his departure from his home country. Correspondence is invited.

b) New students are required to attend Orientation Week prior to the regular registration period. This week has proved an effective means of assisting the student to settle-in to his new environment. Its chief values are believed to be the development of a feeling of belonging and the establishment of friendly personal relationships necessary for effective counseling.

c) All matters pertaining to the life and work of students from overseas are coordinated by the Counselor and his assistants, including:

(1) Utilization of all the administrative, counseling, and instructional services of the College, including.

(a) Counseling by major professor and instructors.

(b) Health counseling, vocational counseling and psychological counseling.

(2) Special counseling of students from overseas During the first session, especially during the first weeks, and decreasingly throughout the student's period of study, the student needs someone to whom he may go for counsel on any matter which disturbs him, whether trivial or important. This requires among other things:

(a) The establishment of a relationship between the counselor and the student which is friendly, permissive, and completely frank.

(b) Authority of the counselor for coordination and for final approval of the program of the student for the first session, including:

(i) Choice of field of specialization

(ii) Choice of major professor.

(iii) Control of amount of work to be undertaken

by the student whose competence in the use of English is limited.

(iv) Requirement of courses or tutorial work in English when needed.

(v) Requirement of an orientation course in American culture and education during the first session to provide experience, interpretation, and counseling necessary for a gradual induction of the student into the corporate life of the college community.

(c) After the completion of the work of the first session the student may return to the counselor for assistance, but by this time he should have become relatively self-sufficient in knowing the resources of the College and in making use of them. He should be more or less on a par with other students, except for special problems peculiar to students from overseas, or to the individual.

III. *Qualifications of the Counselor of Students from Overseas*

I need not discuss the professional and personal qualifications of counselors in general. We are doubtless in substantial agreement that the effective counselor of students needs much more knowledge than most of us have. I assume that the counselor of students from overseas should meet high standards of professional competence. But I am chiefly concerned in this discussion with the personal qualities of persons who counsel students from overseas. No amount of professional preparation and experience can possibly be a substitute for certain necessary personal qualities. Here, I indulge in a bit of idealism.

A. The personal qualities of the counselor must include as large a measure of the following qualities as possible.:

1. Fundamental belief in and commitment to the American way of life; its ideals and aspirations; its problems and solutions; balanced by a belief that the U. S. A. is not perfect, but could be improved and enriched by contact with other cultures.

2. Fundamental belief in and commitment to the interdependence of all peoples and the need for intercultural understanding.

3. Fundamental belief in and commitment to the exchange of students as one of the most effective means of promoting intercultural understanding and *One World*.

4. Fundamental belief in and commitment to counseling as a necessary means of accomplishing the objectives of exchange of students.

5. Genuine sincerity and frankness in all human relationships.

6. Genuine interest and faith in people, as persons, for their present worth and for their potentialities, and wholesome personal relationships with his own countrymen. (The position should not be a channel of escape for persons who cannot get along with their "own people", so seek solace in strangers).

7. Genuine interest and faith in students from other cultures, as persons and as unofficial ambassadors

8. Genuine sympathy with and understanding of the problems of students from overseas.

9. Great patience in assisting students from overseas in solving their problems.

10. Firm belief and confidence in one's self as an effective counselor of students from overseas.

11. Ability to work harmoniously with other agencies and departments both within and outside of his institution.

B. Professional qualifications of the counselor should include as many and as much of the following as possible:

1. Knowledge of the chief characteristics of the culture of the U. S. A.

2. Knowledge of the chief characteristics of other cultures of the world. A genuine interest in learning from students from other cultures is indispensable.

3. Travel and/or study in other countries is highly desirable.

4. Knowledge of the theories and techniques of counseling, of human relationships, and of intercultural under-

standing is essential. If the counselor does not possess such knowledge, he should set about obtaining it as quickly as possible.

C. Necessary administrative arrangements for effective counseling, include:

1. Appointment of a qualified person to serve as counselor.
2. Allotment of adequate time for counseling. It should be understood that effective counseling of students from overseas requires much more time, perhaps twice or three times as much time as that usually allotted to American students

The following schedule is suggested as a target:

- a) 1-25 students, one-tenth of the full time of the counselor.
- b) 26-50 students, one-fifth of the full time of the counselor
- c) 51-75 students, three-tenths of the full time of the counselor.
- d) 76-100 students, two-fifths of the full time of the counselor.
- e) 101-150 students, one-half of the full time of the counselor.
- f) 151-200 students, three-fourths of the full-time of the counselor.
- g) 201-300 students, full time of the counselor.
- h) 301-500 students, full time of one counselor, and one or more assistant counselors.
3. An easily accessible office or student center for counseling and related activities is essential.
4. The counselor must have a liberal schedule of posted office hours and be accessible at other times for emergency problems.

IV. Need for Professional Preparation of Counselors of Students from Overseas

With students from overseas now distributed among over 900 colleges and universities and with over 400 counselors already serving in these institutions, the time has come for special professional education for such service. Such professional preparation should include adequate courses and prac-

tical experience in counseling. While thorough professional preparation as a counselor prior to service is desirable, this is quite impossible at present. The major portion of the counselor's professional preparation, therefore, must be obtained on the job; chiefly through learning by doing in the day-to-day counseling of students, by cooperation with and learning from other counselors, by study of professional literature, and by participation in research in this field.

Among the purposes of the proposed national association of counselors of students from overseas should be the development of professional attitudes, the conduct of research studies, and the stimulation of individuals to seek professional preparation for their work. A first step in offering such professional education will be taken by Teachers College, Columbia University, in the Summer Session of 1948. I quote from the announcement:

Education s209Mx—Practicum for foreign student advisers, 4-8 points. Professor Clarence Linton and Miss Virginia French. 8:30-11:20 a.m. and 2:30-5 p.m., 2 Milbank. Registration by permission of the instructors. Limited to twenty-five students.

This course is planned for persons already appointed to the position of foreign student adviser in a college or university, or other agency; and is designed to assist them in developing more effective advisement and orientation programs for their respective institutions or agencies. Requirements include participation in the varied program in Teachers College; in class discussions of programs, problems, procedures; in readings and research; and in preparation of a comprehensive program for the students' own institution or agency.

In this practicum students from the U. S. A. (i e., counselors of students from overseas) will be full participants along with an approximately equal number of students from overseas in a variety of intercultural activities, including:

1. The orientation course in *American Culture and Education*, where the counselors will serve as consultants and resource persons to help the students from overseas to understand us.
2. The International Film Series in which films of the U. S. A. and of different countries will be used as sources of information and as a stimulus to discussion of the different cultures. In this activity the counselors will serve as informants and consultants for American films and the students from overseas

will serve as informants and consultants for the films on their respective countries to help us to understand them.

3. Visits to schools, to the United Nations, and to other institutions and agencies will provide further opportunities for sharing experiences.

4. Each counselor will be teamed-up with a different student from overseas each week (six different students from six different countries during the Summer Session) for study, conferences, preparation of written work, and other activities. Thus each student, from the U.S.A. and from overseas, will have an opportunity to know fairly well at least six persons from another culture. It is believed that these experiences will encourage genuine sharing of knowledge, problems, attitudes, ideals, and aspirations, from which we should learn something about intercultural understanding.

5. In addition, the counselors will have a two-hour session to themselves with the instructors for lectures, discussions, evaluation of their experience, and for planning programs for their own institutions or agencies.

In Conclusion

The American people's interest in other peoples of the world is running high today. We have come to understand, perhaps only vaguely and superficially but nevertheless with feeling, that our future welfare is tied-up with and dependent upon the welfare of others. Education is teeming with projects designed to teach the techniques of intercultural understanding. Heightened concern with the exchange of students is one manifestation of this interest.

It is the responsibility of educational leaders to provide specific objectives, a sound program, and techniques which will focus this potential of interest and canalize efforts to build a sound foundation for long-view policy and sustained achievement.

I earnestly urge that the ACPA accept the challenge of the opportunity and responsibility of promoting an effective program of exchange of students by:

1. Facilitating the admission and enrollment of students from overseas.

2. Recommending the appointment of qualified persons to serve as counselors to students from overseas.

3. Recommending to your administrative officers and faculty that they support the appeals of the Institute of International Education and those of other groups to impress upon your congressmen and senators the need for adequate appropriations for exchange of students.

4. Recommending that the efforts of your institution be coordinated with those of other institutions and agencies, so that we shall marshal the total force of educational leadership to meet this challenge.

The people of the U. S. A. are investing many billions of dollars in peace. We shall probably invest 17 or more billions to help Europe achieve peace and recovery. This investment is necessary for us as well as for the peoples of Europe and the world. If ERP contributes to the peace of the world it will have been worth many times what it will cost in dollars.

The greatest investment in peace must be in the minds and hearts of people. The program of exchange is such an investment. Now that our government has taken the first significant steps in the G. I. Bill of Rights, in the Fulbright Act, and in the Smith-Mundt Act, we, the college personnel officers of our higher educational institutions, should put our best thinking and efforts into the program.

In closing, I venture a prediction and set a target. We have approximately 20,000 students from overseas today. This is about four times as many as we had in 1938. We shall have 30,000 by 1950 and 60,000 by 1960. This prediction is based on the assumption that peace will be achieved and maintained and the further assumption that our colleges and universities can and will accommodate these increasing numbers of unofficial ambassadors with mutual profit to them and to ourselves. The total cost of 60,000 students from overseas for 40 years would be less than the cost of ERP for the first year.

DEMOCRACY THROUGH STUDENT GOVERNMENT¹

WILLIAM B. WELSH

President, U. S. National Student Association

I SHOULD like to share with you some of the observations that I have made in the last few months while working on the Staff of the National Student Association. One of the important things on which I have been assembling data is—how well is the teaching and practicing of democracy being carried out in our colleges and universities? It has been both encouraging and discouraging at times to get into bull sessions with student leaders from many different types of colleges and to find out what they are really thinking.

We all know that within a university community there are usually three general divisions of personnel directly concerned with the educational process—the administration, the faculty and the student body. It is true that over a period of many years the administration and faculty have been able to develop fairly logical and workable means of coordinating their activities, and the structures by which this can be done fairly easily have been set up. At least it appears this way to the students. But the third group, the student body, has generally been the last group to attempt to coordinate and organize its activities. In fact it would be safe to say that in the majority of cases this has only been going on in the last thirty years, and then with varying degrees of success.

At first glance it might seem that the physical problems of organizing this transient group, which has spent little if any time in the specialized field of higher education, would be insurmountable. However, I believe that surprising results have been obtained in the past, and especially are being brought about in the years since the war.

¹ Discussants whose comments are not recorded are William Birenbaum, Director of Student Forums, University of Chicago; Gordon J. Klopf, Administrative Fellow, Student Personnel Service, University of Wisconsin

Let us call this organized student group that we shall be discussing the student government, and see in what light it is looked upon by the members of the three large divisions on the campus. I have found while talking with faculty members and administrators that some of them hopefully look for some ideal system of student government which can take off their hands the red tape and confusion of student activities. Others of this group seem to think the student government is a fine place to put a great deal of the responsibility for the disciplinary problems that arise on the campus. Others feel that it is dangerous to give students too direct a way to express their opinions and that the student's only role in college is the attending of classes and the passing of exams. And then there is the group who do believe that work in student government is an excellent way of developing potential leadership and understanding of organizational problems, but I'm afraid the the group that looks upon student government as a place to encourage an understanding of the real meaning of living within a democratic society is all too small.

I've also talked with a great number of students, and am sorry to say the general attitude seems to be even more discouraging. All too often I've heard such comments as—"Oh that's just an outfit for the big-wheels to get their name in the paper" or "It's only good to run a dance, but it never does anything for me." Sometimes the students even feel that the real reason for student government is to put up a front for the administration and that really the dean holds a veto over any worthwhile project that might be undertaken.

Of course the whole purpose of outlining these observations for you is to bring out what NSA feels is the real purpose in having a student government. The long range purpose should be almost self-evident: it is to provide students with a practical education in democratic self-government. It is in effect an insurance policy for a nation that prides itself on the capacity of its people to govern themselves. We feel that in any discussion of this matter student government should be regarded primarily as a learning process in which the lesson taught through actual practice is the hard and tiring way of an ordered democratic society. We recognize that democracy is probably

the most difficult form of government and system of ordering a society ever conceived by man, but that it is the only successful one in which the individual is given a full opportunity for intellectual and moral growth and in which he may use his full capacity to contribute to the common good. And within this context we feel student government has an important role to play. As the report by the Harvard committee on *General Education in a Free Society* expresses it "Student government is valuable in shaping the quality of later citizenship. It is only when the student faces the actual difficulties of governing by democracy that he begins to appreciate the complexity of a free society. To learn to resist pressure, to discover the power of a minority, to have free speech used against one, to prescribe rules and then to abide by them is training of the first order." This then is what students may learn if they have the opportunity of participating in student government activities.

The student government itself, like any government, has two major functions to serve the people it is made up of, and to represent them democratically to the other groups with which those people must work and live. Now to discover whether or not student governments really have the chance to do these things, and the opportunity to learn those things listed by the Harvard Committee, some system of evaluation should be developed. The individual college should be able to evaluate the workings of democracy within itself, and then the level of evaluation of democracy should be brought down to the individual student. We must have the tools for analysis within the reach of the campus. It is very commendable to talk of saving the remnants of democracy on the international scene, or in Washington. But the evaluation of what I am looking for must not begin there. It must start within the fraternity house and in the Union building. There must be some new sort of check list or criteria of action by which every student can evaluate his participation in the democracy of the campus.

It will do little good for NSA to tell a student government where it is making mistakes or for us to write an editorial in the *NSA News* about the undemocratic student government at a specific school. It will do little good for members of the

student government to call the members of the student body apathetic or to criticize the members for taking little interest. Each individual student government must work out for itself the techniques of making democracy a learning process, and each student must be urged to participate and then to evaluate his own part.

I would further like to see some method of evaluating democracy not only within the student government, but within the whole educational community. I'm afraid too often it is the bad example set by the faculty and administration of a college that puts our students on the wrong track when they form a student government. It is much easier to sit behind a shiny desk and say to a group of students that this can be done and that can't. It's easier for the president of the student government to run his affairs with a yes or no answer, but we would all cry loudly if this were the method used in our local or national governments. It is always harder to reach agreement through the free exchange of ideas and by majority decision.

I hope that from this group will come the suggestion to each of your student groups that they wind up their activities this year with an overall evaluation of how well they were able to use the democratic process in their work this past year. Before they can do this, of course, they will have to evolve some sort of criteria to judge the democracy of their work. I hope that many of them will take the time to do this. The results of this should certainly be kept for the new members the following year, so that they may see the strong and weak points of their predecessors. Then each of us in our work with student governments should stop and take a look around. Do we feel that discipline is the prime objective of our work with student activities? Are we afraid of making a mistake by letting the student government stumble slowly along in the democratic process?

The example that each person in the field sets in his own practices will be a determining factor in how student groups function. Let's not wave the American flag and think it is grand to give lip service to democracy, but rather let us think out how we can put democracy to practice within our own

spheres of influence, and make it live and grow. If democracy fails within the educational circles of America, I hate to think how miserably it will fail in the city, state and national spheres.

Perhaps I have painted a rather bleak scene. For a minute I would like to think of the reasons I believe the university community is an ideal place for even the transient group of students to learn the workings of democracy. Here the student has an opportunity to deal with problems within his own scope without too much outside interference, and in areas where mistakes can be accepted. Too often it is because of fear of public opinion against the university that sufficient areas of defined authority are not turned over to students--and because mistakes might be made. I feel that one of our first jobs is to inform the public that student government is a learning process, and is an accepted part of each university system. Also working in a student government gives the students a chance to see that the solution to organizational problems does not lie in merely writing a beautiful constitution. I don't know how many times this year I have heard that some student government is rewriting its Constitution. I wish I could call out to all of them not to get jammed-up in words, but to put their efforts into making what they have work.

Let us look briefly at some of the *specific mistakes* that are being made today in this field.

1. I feel that there has been a failure to outline clearly the areas of authority in which a student government may work. I realize that in every instance this authority will properly come from the administration of the university, but it must be clearly outlined and understood, and it must be complete. The advisors that work with student government should realize this and constantly be on guard to see that they do not try to pressure decisions in the areas in which authority has been clearly delegated.

2. Highly organized campuses, especially those with independent and fraternity groups, are prone to take the politics of student government as something that can be kicked around for excitement. This talking and playing politics for politics sake is often exciting, but if that is the end, it is usually the end of functional student government.

3. Many of the larger university student governments find

their meetings almost completely dominated by consideration of athletics. In some cases this has been carried to the extreme and resulted in the complete neglect of broader considerations of student welfare.

4. One of the worst faults that I have run into is the student government's accepting the responsibility for a conglomeration of disciplinary powers over student behavior. I am sure that wherever discipline is the stated or actual prime purpose of having a student government, the student government is not very effective in that field or any other. Too often I have found that rules and regulations are left unexplained to new generations of students, and the only people who know that discipline is in the hands of the student government are the administration and a few of the student council members themselves. Disciplinary functions of student government should be taken on only as the whole student body is willing to accept the responsibility. The ideal situation might be for a set of jurisdictional by-laws that required reacceptance each spring by the student body. This would always keep the responsibility before the changing student body, and give the student council a vote of confidence in this very important field. A group that starts its existence with an imposing list of jurisdictional or disciplinary powers won't have a very long life. It will soon find that it is losing the respect of both the administration and the student body as it finds it can not use police powers wisely. If it becomes clear that one or two disciplinary powers are being handled effectively by the student government, an additional power might be added each year until a workable list is obtained. The key to the ideal situation, however, is to have the whole student body accept the responsibility, and not just to have the executive body of the student government act as a policing agency.

Now let us consider where student government does function successfully, and in what areas *constructive projects* that are beneficial to the campus are being carried out

1. The coordination of campus social activities, and their expansion until every student has an opportunity to participate, has been the concern of many student governments. Several have expanded intramural sports.

2. Some students governments have surveyed housing

facilities in the surrounding community, and checked up on local restaurants to see that they are keeping high health standards.

3. Many of the student governments have taken on the project of running a campus community chest and consolidating all of the varied relief and welfare drives into this one campaign.

4. There has been increasing interest in the problem of constructive student evaluation of teachers, and several student governments have taken the lead in coordinating this project with faculty groups.

5. At one school the student government has been instrumental in obtaining a place for at least one student representative on every university committee, thus tying in more closely the loose ends of the educational community.

NSA's role in this broad picture of student government is primarily to make more effective the work of the student body in the university community, and to act as their representative and consultant in national and international educational, and governmental groups.

One of our first jobs on the national scene this year was to compile material and information on the actual structure and functioning of student government. We have published a booklet on *Student Government and Leadership in Higher Education*-- and along with it sent plans for student government clinics that can be held by interested groups on a single campus, and by several schools in a region. A survey of 360 colleges is being completed to show the functioning and structure of student government in all types and sizes of colleges and universities in the United States.

Several of our sub-commissions on various campuses are doing similar planning and survey work in such fields as seating in football stadiums, use of upperclassmen as advisors, personnel and guidance, orientation programs, election procedures and campus social activity organization. Two of the regions are working on regional cultural festivals, and inter-school dramatic and musical events have been initiated and coordinated through the NSA channels.

Some of our most immediately successful work has been done

in the field of international student activities. Under the sponsorship of the NSA the MIT student government has helped to arrange for the study of 80 foreign students, expenses paid, at a summer course at their university next summer. Plans are under way with several colleges for a program developed jointly by the Canadian Federation of Students and the NSA for exchange of students each year between the two countries. The student governments at many schools have actively supported the World Student Service Fund drives for the first time. NSA is a national sponsor of VSSF.

NSA has made special arrangements for 100 students to take a planned tour of Great Britain, Holland, Belgium, and France this summer. All arrangements were made through the various National Unions of Students in those countries, and we have supplied each student government with booklets and bulletins giving a full picture of all the study, travel, and work opportunities abroad this summer. Student governments are beginning to make these services available to the members of their student bodies.

We hope that through NSA we can help individual students to look beyond their classrooms, and to encourage them to help themselves with their educational problems. NSA itself is a learning process and will never reach any ultimate goal in itself, but it will make the job easier for its member student governments and will provide them with channels through which their activities can be expanded into the regional and national scene. It is with this in mind that students from all types of colleges are willing to cooperate and to work with each other in the National Student Association.

HUMAN NATURE AND EDUCATION

HAROLD TAYLOR

President, Sarah Lawrence College

It is with some diffidence that I speak this evening about a subject as complicated as human nature and human growth before an audience so knowledgeable and so skilled in both of these fields. I am a philosopher who gave up thinking to become a college president. I feel that I might have some professional competence, remaining from that happy life of the past, to deal with questions such as the structure of the universe, the destiny of the human race, and the nature of freedom. But as to the bigger questions of giving advice to college students about what to study, how to study, when to study, how to know what one can do, how to get a job, and how to avoid going mad, which I take it is the business of the dean and the personnel people, these are much too intense and difficult matters for me to handle with any confidence. What I shall do, therefore, is to make philosophical remarks about children and how they grow, then go on to recommend a philosophy of education to which I have become attached, and end by suggesting one or two revolutions we might start which I think would improve matters considerably.

First of all, I would like to say something about human nature, and the qualities of human nature which we look for in liberally educated people. I believe that the liberally educated person is one who is deeply interested in life and who enjoys it one way or another, a person who is sympathetic and generous in his attitude to other people, to other cultures, and to other countries than his own, who accepts his world and himself as a growing, changing enterprise in whose advance he has a significant part to play, who is sensitive to the beautiful and the ugly in actions and objects, who believes in human rights and human freedom, who has a degree of knowledge and knows how to get the knowledge he doesn't have, and has at least a

moderate skill in the art of living. It would also help if he were rational.

There are a good many ways in which these qualities are learned, and a great many different kinds of people who show them, - composers, poets, scientists, farmers, workers, and sometimes even professors. It is not necessarily a matter of school and college education, since the qualities develop in other places. It is a matter of educating the feelings, the thinking, and the values of each person.

It is clear, I think, that values are learned, not always consciously, by the particular set of situations in which people spend most of their time, by the direct and indirect personal influence of parents, teachers, friends, and employers. I believe I could defend successfully the view that the value of things and ideas is learned only by the immersion of the individual in the stream of human relations which make up his daily life. The value of money is not learned through the study of economics but by the immediate awareness of what people can or cannot buy with it. The value of democracy is not learned by the study of American history or Greek philosophy, but by learning what it actually feels like to be democratic. The value of ideas is learned by working with people who have them, who are willing and able to exchange them.

To me this means that a truly liberal education begins with the child when he explores his talents and desires. His mind and the rest of him all grow together as he plays with clay, rides a tricycle, wrestles with his contemporaries, and learns to cooperate with his friends. For the child, fortunately, it is almost impossible to separate the training of the mind from the other things he does, since it is very clear that his mind is that part of his life which helps him to explore his world successfully and to satisfy his own need for knowledge and understanding. A genuine sense of wonder exists in each child, and each one asks about the origin of things and why they happen as they do. The child continually looks for values and for truth, and his search leads him to the things which later make up the formal studies. The way in which we measure

the intelligence and talents of children when we are not involved in the inhibiting processes of what we call their education, is by the way they respond to the situations in which we put them, the liveliness of their interest, the readiness with which they grasp the meaning of what we say and what they say to each other. This of course involves the things they know, but the things they know are those they have learned through living them. In children, we recognize the fact that there is a close relation between personality and intelligence, and when we speak of children, we try to assess them as creatures who are more or less pleasant to have around, more or less able to cope with themselves and with others, and more or less worthy of any efforts we may wish to put into their further education.

Nor do we ask constantly for tests of memory, or prescriptions of subject matter, since in the young, life is a whole and cannot yet be sliced into academic segments. We are glad to have the child involved and growing to his maturity by a variety of means. We want him to learn to live in his world by seeing relations between one thing and another (philosophy), understanding how to deal with himself and other people (psychology), what are the names and places of things (science), how to enjoy himself (the arts), how to exert influence on his parents and other people (social science), how to count and work things out (mathematics), how to imagine people and situations (literature), and what are the most desirable ways to behave (the humanities).

Altogether, if we are parents or teachers, we want the child to change and grow in whatever ways seem most significant to his talents. Our role is to set the moral structure widely enough to allow freedom, and narrowly enough to prevent chaos. We do not want each child to be the same, we usually want him to be like his mother or his father, or failing these, like an important person outside the family whom we respect and admire. In former years, this was invariably a captain of industry, or the head of something important like a bank. Now we are for the most part content with less. We will usually settle for an offspring who seems to have himself in hand and who is not too severe with his parents.

I believe that the diversity of human nature is one of its chief attractions, and that any effort to work against the spontaneous diversity of human talent is doomed to failure. It will fail either by accomplishing its ends, and making a standardized product of dull uniformity,—and thus something less than human, or by producing an interesting rebel who refuses to be standardized. In either case it fails. My suggestion is, therefore, that as far as human nature is concerned, we put to full use the things we know about human behavior from studying the young. This means that education will have to stop acting as if human nature is everywhere the same, will have to deal with people one at a time, and will take as its goal to release in each person the talent which lies within.

I believe that the philosophy of education dominant in America today is the opposite of this and is calculated to produce a standard college graduate, approved by the authorities, possessing a standard body of knowledge, and guaranteed to be educationally house-trained. During the past five years, there has been a great amount of talk throughout the country about reform of education. On the whole, this has resulted in a move backwards towards the tighter organization of the intellectual life of the college student, and a reaction against new ideas for educational change. Educators, I am afraid, are now thinking more about mental training and the faculties of the mind than about developing mature persons. The reasons for this reaction are in large part the result of the failure of the elective system to give genuine unity of knowledge in education.

The unity now sought is in the direction of general education, and is the result of a sincere effort to avoid the hazards of special and narrow education, and at the same time to achieve a breadth in knowledge which will equip each student for contemporary life.

However, when educators have set to work to carry out reforms, they have composed their committees, and their reports about committees, around the idea that the whole matter is one of subjects to be studied. Professors from each department in the university have tried to think of subjects they

wish they had been compelled to study when they were young, or subjects which, by sheer number of departmental members, have been voted into prominence in the curriculum. Having thus decided what these subjects are, a new and required curriculum has been made, intended to give a general view of everything, or, to put the aim another way, to make people specialists in the general.

In all this manipulation of subject matter, the student, his needs, his world, and his judgments about his needs, have been ignored and excluded. Along with the student have been excluded the dean of students, the psychologists, the counselors, the ones who know what bothers the boys and girls. They know because they are the ones who talk to them outside of the classroom about things of deepest significance in their personal lives. In the absence of the student and his spokesmen, the counselors, the professors have talked and voted themselves into a standard curriculum which combines and rearranges in new clusters all the subjects and courses which have been given harmlessly enough for years under different and less complicated titles.

I believe that the clue of these ambitious efforts at intellectual purity is to be found in the classical view of human nature held, a good deal of the time unconsciously, by members of the academic profession. According to the classical view, the individual human being is identical throughout the whole of the species, he contains a mind in the same way that an egg contains a yolk. Just as there is a universal method for hatching eggs by placing on the nest the curriculum of the universal hen, so there is the classical curriculum, everywhere the same, which can guarantee the hatching of correct knowledge in the universal student. To use language which is more respectable, the mind of the individual, in the classical theory, is a rational entity, constant throughout its human occurrence. Although the body may change, and may vary from person to person, and may occasionally get the mind into trouble, it has only incidental relations with the mind, and thus has little to do with liberal education, or the intellectual life. A major part of traditional education, for this reason, consists in wrestling with the body and trying to keep it from interfering with social and intellectual processes.

The flaw in this view of human nature is that the things we really care about have their roots deep in the body, and that desires, ideals, sex, sensibility, joy, and a number of other useful assets, are more basic to human life than the incidents of the intellect. It is important to add immediately that the intellect, or mind, is a great aid to us in achieving the ends towards which our desires would have us move. The intellect also gives us considerable help in deciding in which direction we wish our desires to set about moving us.

A modern theory of human nature and education is therefore more complicated than the classical theory. It accepts, for example, the influence of social environment on the personal character of the individual, and works by all the ways it can to discover what changes occur in the one due to changes in the other. It accepts the existence of an unconscious mind whose effect is perhaps more powerful in some instances than the conscious, since it operates to influence desires when we are unaware of its very existence. A modern theory assumes that there is a direct relationship between childhood experience and character in later life. It accepts the view that motivations and talents are different from person to person, and different in each person at different times in his life. In other words, it accepts as observably true the fact that the individual human being is a special case, and that if education is to be effective, it must deal with the emotional, intellectual, and social needs of the people it is serving. In a very real sense, the whole of education is guidance to life.

Those of us who are entrusted with the guidance of the young toward a full and happy life are placed in a situation of crucial importance. I think it is clear that we are living in a period of social tension and confusion of purpose which has no precedent in the history of civilization. Although we may not feel that we have the knowledge and the insight to tell our students how to act and how to live so that contemporary civilization is aided to survive, it is our duty to make the effort, both by becoming as clear as we can as to the nature of our crisis, and as clear as we can about the best means to overcome it.

The elements of social crisis have no doubt always been present during the history of the last two thousand years, yet in no age before our own has man been so self-conscious about

it. We have Mr. Toynbee's word that when civilizations appear to be at their most healthy, prosperous, and blooming, they are actually rotting away at the core, and that the ignorance of one's own precarious position is simply a failure to recognize social crisis until after it has happened. Thus, the rosiest apple contains the biggest worm. Since Mr. Toynbee's hypothesis cannot be verified, and is much too disquieting to use in thinking about one's own personal condition, I suggest reversing his view and, by a happier logic, which has just as much chance of being true, hold that when civilization seems to be at its worst stage of crisis, it is actually at its most promising. In this way, I find that I can face the future very nicely and with an amount of historical dignity equal to that of Mr. Toynbee.

I believe that it is important to recognize that our present crisis has special qualities, and that it is not just another period of the sort which has followed wars in the past. Its prime fact is that for the first time, we have the means, and to some degree the will, to destroy the whole of contemporary civilization. Our age is marked by its use of power, violence, terror, and intimidation as the standard means of settling social conflicts, with a sharp decline in the use of persuasion, appeal to moral values, and respect for personal rights. The symbol of the age becomes Ghandi, the world's only apostle of love, assassinated by an intellectual political opponent. Ten years ago, the symbol of the Western crisis was Chamberlain, his umbrella, his gout. The age is now marked by an enormous increase in the sources of various powers useful for bullying people, whether through the movies, the radio, the newspapers, atomic energy, or economic monopolies. Its favorite words are confusion, action, and tension. It is marked by the growth of the importance of politics and collective thinking and acting for the benefit of states. But perhaps more significant than any of these things, and in large part as a result of them, it is marked by a disintegration of moral confidence and conviction, and the disappearance of the individual.

The function of education therefore is to restore to the individual his confidence in himself and his ability to solve his social problems. The prime educational need is for knowledge

of our present situation, and an understanding of the human forces which are making our future. Otherwise there will be no future, but only an uneasy five years with an enormous bang at the end of it

I have read recently the statement, written by an educator from this vicinity, that in order to survive, modern man needs only to know the content of a group of special books. It has been said that already the discussion of these books has altered "the tone and subject matter of conversation in considerable areas of the United States." This may be true. The drawing rooms of America may now resound with the ring of famous names, but the society which surrounds them remains largely unaffected. I have seen a classification of one hundred ideas which contain in themselves all the central concepts necessary for dealing with this or any other crisis. Such a group of classified abstractions is of the same order of meaning as the Chicago telephone book. If you work at them long enough you will find a human being at the other end.

Fortunately, we have a clear line of action shown to us by those American educators who have written the *Report of the President's Commission on Higher Education*. A modern philosophy contained in that document declares the action should be to educate the individual, wherever he may be, to give to each the kind and amount of education and training appropriate to his talents, to give to each an understanding of the present, and to give to each a set of ideals towards which a better life and a better society can move. These ideals are learned and absorbed by so organizing the college life of each student that he shares responsibility for the lives of others, has a chance to think and act independently, and is able to receive the advice and aid of his teachers and counselors

In other words, the whole of college life should be made into one piece. The gap between the academic and the personal must be closed by the union of counseling with learning, learning with acting, and acting with knowing. Here is the significant place for the educators whose assigned function it is to deal with the social and personal life of the student. In their hands lies the opportunity to teach democracy and with it intellectual, emotional, and political maturity.

Admissions programs can be devised to select students most qualified in personality and intelligence, to remove racial barriers, to bring in diversity of talent. Testing programs can be revised to eliminate the statistical approach to human values. Housing arrangements can be made to eliminate racial discrimination, and where such discrimination appears on our campuses, we can invent educational means of correcting the attitudes of those who practice it. It is not enough to put in a required course in race relations for white Gentiles and to sit tight. It is a question of blending the diversity of students from all social, racial, and economic groups into a community of free-thinking, humane, and generous young men and women. Social snobbery, bred in the fraternities and sororities and fostered in campus politics, can be dealt with by deliberate educational policies developed by students and faculty. Anxieties about jobs and the future after college can be turned into purposeful educational planning when vocational guidance is conceived as the attempt to develop intellectual and social consciousness. Political maturity can be aided by encouragement of free discussion of all ideas, radical and conservative, and of all the issues currently facing American youth. Simply to take, one at a time, the issues raised by the coming elections, to enable each student to discover for whom he will vote, is a means to understanding and an aid to wisdom.

The revolution I recommend can therefore be stated very simply. Let's give the colleges back to the students. The meaning of education can only be grasped by those young people who seriously take hold of their own lives and make their own knowledge and value systems. When academic educators attempt to take over this task for the benefit of the young, it simply establishes an illusion of control by the few over the many. The influence of the classroom as an educative factor sinks to a minimum and the growth of the young self occurs without benefit of intellectual guidance. Only a life in the college community in which each student feels that his presence makes a difference, and in which he feels he has a personal stake in deciding questions concerning his own education and college policies, can give to our young the maturity

they are willing and anxious to gain. This is particularly true in our present situation of excess numbers and standard programs. The individual is lost in a welter of requirements and people. To restore him to the center of the institutional life is the basic need and the central purpose both of professors and deans. At present it is not being done by the professors. I am suggesting that college students are mature enough to be trusted, wise enough to be guided, and eager enough to be educated. Let us therefore give them more responsibility and more chance to work out their own destinies and education.

The new curriculum for American education must therefore extend beyond the classrooms, the libraries, the laboratories, and the offices, deep into the elements of contemporary life. It must touch the personal attitudes of the student at points where he can feel his own responsibility for thinking and acting well. The impact of that curriculum will not be confined to the subject matter of courses. It will spread through the whole of the college. Here the test of whether a subject is part of a genuinely liberal education will not be its traditional status in college catalogues, but whether or not it yields an understanding of the most crucial questions of our present lives.

Finally, the curriculum must, throughout the entire scope of its action, be concerned with art. The joy of creating one's own knowledge from a diversity of sources is one kind of art. The joy of making and appreciating the objects of the artist, is another kind. But as in the case of the child, of whose philosophy I have spoken wistfully earlier this evening, the joy of living, growing, and understanding is the basic art to master. The value for which we search is the one we can gain only by putting together a life which has its uniquely private rewards and makes uniquely its creative contribution to the process of society. If we believe deeply enough in the search, we can, by our own future actions, show that human ideals are greater than the material and social embodiment by which they are presently expressed.

SOME IMPLICATIONS OF CLIENT-CENTERED COUNSELING FOR COLLEGE PERSONNEL WORK¹

CARL R. ROGERS

Professor of Psychology, University of Chicago

THE attempt to utilize the principles of client-centered counseling is in one respect an irritating thing. The basic hypothesis appears to operate with sufficient effectiveness in therapeutic work with maladjusted individuals, that one is continually plagued by the question "If it works in dealing with the person with problems, might the hypothesis prove true in this other type of human relationship, or this, or this?" So today I shall pass on to you some of those questions that have occurred to me and perplexed me in regard to various phases of student personnel work at the college level. I hope that these questions will provide a basis for discussion.

I do not wish to talk about client centered therapy as such. I am sure many of you are familiar with some of the attitudes and ideas related to that point of view. Perhaps as a point of departure I might try to state as adequately as I can in one sentence the hypothesis which is central to this orientation, a hypothesis which appears to find backing in our clinical experience and in the slowly growing body of research data. It could be stated in these words: that the individual has a high degree of capacity for making an adjustment to life and its problems, a capacity which can be most fully released in a non-threatening psychological atmosphere which is permissive, accepting, and deeply understanding, where the individual is able to consider himself and his situation clearly, including those elements ordinarily denied to conscious awareness. This has been the

¹ Discussants whose comments are not recorded are Edward S. Jones, Director of Personnel Research, University of Buffalo; Esther Lloyd-Jones, Professor of Education, Teachers College, Columbia University; James A. McClintock, Director of Personnel, Brothers College, Drew University; Willard W. Blaeser, Dean of Students, Washington State College.

hypothesis underlying our clinical work, and it is from the orientation thus provided that I should like to raise certain questions about college personnel work.

I

The first question is, "Where is knowledge of the student integrated as a basis for action?" As one studies college personnel work, the heart of it appears to be knowledge. There is knowledge of the student's intellectual abilities, his vocational aptitudes, his personality structure, his emotionalized attitudes, his academic history, his past and present grade record, his family environment, his physical and neurological condition, his dormitory behavior, and so on and on. The various ways of obtaining this knowledge have all been carefully considered and developed by workers in the field. But it is obvious that none of us is interested simply in the accumulation of knowledge. It is knowledge as a basis for integrated understanding, planning, and action, which is important. And there has been surprisingly little thought given to the question, where shall we lodge this responsible integration of knowledge, this function of responsible planning and action?

There is implicit in much of the literature regarding student personnel work the concept that this responsible function should be lodged in some designated individual such as the counselor or the dean of students. Sometimes it is implied that responsibility should lie with a group such as a case conference, in which both the knowledge and responsibility are pooled. At the other end of the continuum is the less accepted concept that the integration of significant knowledge, the responsible planning of action, the taking of action, should be centered in the student himself. The locus of the function of evaluation, according to this point of view, should remain in the student. The increased interest in using college program-planning as a responsible learning experience for the student, instead of an administrative and guiding function, is one evidence of this point of view. Another is the attempt to have the student select the types of psychological tests he wishes to take, and to have him also bear a primary responsibility for the meaning the test results will have for him. Thus there are a number of

people in the student personnel field who are giving up the notion of thinking *about* the student and *for* the student, in favor of thinking *with* the student

The question I am raising may be put in another form. What are the long range implications of a theory of student personnel work which puts the integrating focus of personnel information outside of the student himself? Speaking from the point of view of my own clinical experience, it appears to me that effective integration of knowledge is best achieved within the student, and that this has concomitant values in developing independence and maturity. I tend to favor this end of the continuum not only because its effectiveness appears to be borne out by clinical experience, and by some research, but also because I happen to favor a deeply democratic political philosophy. I feel that any point of view which tends unnecessarily to substitute direction by others for intelligent self-direction, leads away from democracy.

One of the things which frightens me about student personnel work is its complacency in regard to its long range effects. If we believe that time will tend to perfect our tools and extend the functioning of the personnel point of view, then we believe that it will eventually be applied to us. If we are happy about its application to students, then we will be happy when it is applied to us. But I doubt this. Do we believe that we will be better, more effective faculty members when the dean, or a dean of faculty personnel, has the complete record of our lives, our abilities, our personalities—the same type of record we are striving to build up in regard to students? When your request for curriculum changes, for better teaching facilities, your request for promotion, your difficulties with the Buildings and Grounds Department, your tensions with your wife, are all understood by a benevolent dean in the light of your test results and your Rorschach pattern, will you, will your university be a better place? Certainly much wiser guidance can be given you. You can be helped to see that promotion is hardly justified in view of your ability on the College Faculty Aptitude Test, which has enabled us to select such a homogeneous faculty. It can be gently hinted to you that the dissension in your home and your truculence in faculty meetings as well as

your disputes with the janitor, grow out of your unresolved conflict in regard to authority, and that psychiatric guidance would assist you in all three respects. When you leave for another university, your cumulative record can be forwarded to your new dean, who can also evaluate you and your behavior in the light of all this knowledge.

Do we like that picture? It is the direction in which an increasing portion of the world is moving. Life seems too complex. We should surrender the responsibility for our lives into the hands of someone wiser -- a dean, or a psychiatrist, or eventually of course, into the hands of the state. If this is a social trend, then should we not go along with it? Should we not be happy that much of student personnel work is already in tune with this point of view? This is a basic question which I hope the panel will discuss. I will simply state that for myself, the only reason for opposing the trend is that clinically I find it to be true that though an individual may remain dependent because he has always been so, or may drift into dependence without realizing what is happening, or may temporarily wish to be dependent because his situation appears desperate, I have yet to find the individual who, when he examines himself and his situation deeply, deliberately chooses dependence, chooses to have the integrated direction of himself undertaken by another.

Though I have described the issue all too briefly, and may have confused rather than clarified by looking at it from various angles, this is the deepest question which I should like to raise for discussion by the panel. I will repeat it in a somewhat different form. How do we evaluate the tendency in student personnel work to build up an increasingly complex body of information about the individual, the organizing center of which is located outside the student himself?

II

There is a second general question of quite a different sort which I should also like to raise. It is this: "Can the basic attitudes which appear to be effective in individual counseling of a client-centered sort, be applied to work with groups of stu-

dents, to work in committees, to the teaching of classes, to the administration of student personnel staff groups?"

The experience we have had in group therapy and in dealing with group dynamics is not as extensive as our experience in individual therapy. Research in this area is still relatively meager. Nevertheless certain statements would appear to be justified, which might be descriptive of our clinical experience to date.

When a group faces a problem or problems;
 When the leader is genuinely free of a desire to control the outcome of the group experience,
 When the leader respects the capacity of the group to meet its own problems;
 When the leader has skills in releasing individual expression;
Then
 The individuals begin to express attitudes, and to be responsible for their own attitudes;
 Both conscious and heretofore denied attitudes are brought into the discussions;
 The individuals perceive their attitudes and themselves differently, and consequently perceive their problems differently.
 The group develops a strong "morale", and feeling of closeness;
 The individual or the group depending upon the type of problem works out constructive ways of realistically handling the situation.

Could these group procedures be applied to situations faced by student personnel workers? There is reason to feel that they could. Cautious experimentation would seem justified. We have handled groups of students with personal problems and tensions. Results vary from the student who feels it has been an interesting experience of no particular personal value, to the individual who feels that group therapy has been one of the most valuable experiences of his life, and whose behavioral changes bear out this judgment. We have also tried group therapy as a means of reducing anxiety prior to comprehensive examinations. The results are not conclusive, but warrant further exploration.

A situation faced by the student personnel worker in which such methods might be used, is the task of freshman orientation. Often the orientation process consists of a week of meetings in which information is gently and steadily poured over the head

of the student, with wienie roasts and inspirational talks to make the process bearable. Suppose that instead beginning students met once a week during their first semester in groups of not over twenty, to talk out their confusions, their fears, their resentments, their feelings of inadequacy, and their sense of growth. The group could be conducted in such a way as to make it *their* group, an opportunity to develop a clear and realistic orientation to their own situations. Information, probably mostly in written form, could be made available to them. There is reason to believe that students would use such an opportunity in a responsible fashion to work out a personal orientation to the college situation. Group morale appears to be an almost inevitable accompaniment. I know of only one situation in which just such an orientation procedure was used, but it was regarded as highly successful.

Let us examine another area in which personnel workers are involved, classroom teaching. As we have attempted to apply learner-centered procedures to university courses, we have become convinced that it is difficult, and that we have not solved all the issues. But it is also very clear that such an approach can have extremely vital results. It means largely giving up the attempt to *teach*, and attempting instead to create conditions which facilitate *learning*. This is a very different thing. When it is achieved, when the leader is genuinely a catalyst and not a mentor, the group shows remarkable capacities for coming to grips with the real issues, for self-initiated learning, for intellectual and emotional growth. As one student said at the end of such a course: "All my life I've made a ceremony of burning my notes at the end of a course, to show I was finished with it; in this course I have a totally different feeling, that I have just begun to learn, and that I want to go on."

The same policies would apply to our handling of student or student-faculty committees. If we clearly leave the responsibility with the group, if we are genuinely understanding and nonjudgmental when emotionalized attitudes are expressed, we are likely to find vigor and life and the taking of responsibility by the group. Many college personnel workers have realized this and have come to utilize more and more deeply

the strength of the group, rather than to try to lead through the imposition of their own strength.

Let us look too at the way in which we function as administrators. Many of us have some responsibility for people under us: staff members, clerical workers, and the like. Does the client-centered point of view have any implications for the way in which we carry on such a function?

Here I should like to speak quite personally of my own experiences. Since 1939 for nearly twenty years I have had administrative responsibility for staff groups of one sort or another. I had developed ways of handling administrative problems—ways which had become fairly well fixed. Certainly as I became more and more deeply interested in a client-centered type of counseling, it was furthest from my mind that it would ever affect the way in which I dealt with organizational problems. It is only in the last two or three years that I have been really aware of the revolution in administrative procedure which it might bring about. I would mention again a point I made at the outset, that the effectiveness of a client-centered approach in counseling means that these concepts continually force themselves into other areas where one had not thought of using them.

For myself, I have found it both difficult and rewarding to attempt to apply these concepts in administration. I am sure there are many aspects of the application which I do not understand, or do not see clearly. Yet there are certain questions which I can ask myself about the administration of an organization, to which I believe I see the answers. I shall list and comment on these questions.

1. Do I trust the capacities of the group, and of the individuals in the group, to meet the problems with which we are faced, or do I basically trust only myself? I find that when I take the risk, the gamble of resting my confidence in the group, ingenuity, responsibility, and strength, are multiplied. If I am fearful of doing that, and rely on myself, this produces in the group passivity, a willingness to sit back and criticize, and dries up initiative and constructive effort.

2. Do I free the group for creative discussion by being willing to

understand, accept and respect *all* attitudes, or do I find myself trying subtly to manipulate group discussion so that it comes out my way? I find that this tests one's basic philosophy very deeply. It is all too easy to give a group pseudo-freedom, just as there is much pseudo-client-centered counseling going on, in which the atmosphere is permissive up to a point, but then controlling. The results are the same in counseling or administration. When there is a genuine willingness for all attitudes to be expressed—critical and hostile as well as constructive—then the group senses the fact that it is their organization, and they respond with vigor, with loyalty, and with responsibility. When the clerical staff is as free to contribute attitudes as the professional staff, then perhaps this principle is most deeply operative. On the other hand, if the freedom is of the "pseudo" variety, then suspicion develops.

3. Do I, as leader, participate by honest expression of my own attitudes but without trying to control the attitudes of others? The essence of a client-centered point of view is not passivity, as so many have supposed, but respect for the right of the person to be himself. In a situation such as a staff group, where I am most certainly ego-involved, it is as important that I express my feeling as that the next person does. But this will again test deeply the leader's philosophy. He can express his attitude in such a way as to imply "and you had better think the same," or his expression can on the other hand imply, "this is just one feeling, and others may have very different attitudes."

4. Do I rely upon basic attitudes for motivation, or do I think surface procedures motivate behavior? It has been our experience that when a problem is felt by the group, and freely and openly considered, and a way of meeting it is discovered and experienced by the group, action along those lines follows. If this process has not been achieved, no amount of formal agreement will bring constructive action. Motions, rules, formal decisions are as meaningless as New Year's Resolutions for the individual. In our Counseling Center staff, formal motions are almost unheard of. The group seems to be an organism, and when it feels itself to be clearly integrated, action

follows inevitably. When it is in conflict, action is confused or conflicting, and no amount of neatly typed policy will make it otherwise.

5. Am I willing to be responsible for those aspects of action which the group has delegated to me? If I do not wish the responsibility I should say so. If I accept it, I am obliged to carry it out.

6. Do I trust the individual to do his job? Here we plunge directly into the question of what is supervision. If supervision is the task of an overseer, directing the individual as to how he does his work, then I think much of what I have described is negated. We have come to put new meaning into the term supervision. We regard supervisory contact as the opportunity which the individual has to think through more clearly the problems he is meeting on his job—the unpleasant demands made upon him, and the way he will adjust to those; the failure experiences he is having in his counseling; the problems he may feel in his personal orientation to the staff. We find that the more the individual is given the responsibility for his job, the more deeply he accepts it, but when someone else assumes that responsibility then his attitude is, "I just work here."

7. When tensions occur, do I try to make it possible for them to be brought out into the open? I think administrators tend to think they are doing well if no tensions are evident. On the basis of our experience it seems much sounder to accept the fact of tension as basic, and to learn to deal with it. People, because they are living people, are bound to be at times dissatisfied, to feel out of line with the group, to feel jealous, to feel critical of others, etc. We have come to believe that it is only when a tension is displaced in its object that it is dangerous. If the staff feels I have been too dictatorial or have played favorites in some action of mine, only constructive thinking accrues if that is told to me. The experience may be painful, but it results in growth. But if it is bottled up, and expressed only in opposition to some new policy proposal, then it does not dissolve, but tends to become heightened. If staff member A is resentful of staff member B whom he feels to be too ambitious, the expression of that feeling to the person toward

whom it is directed is again, painful but constructive. But when this jealousy or hostility is expressed by A only in criticizing B to others, or only in opposing B's ideas about counseling, then tension grows. So we have found it highly important to try to create an atmosphere in which real attitudes can be expressed toward their real objects. When this occurs tension is reduced, and almost inevitably perceptions (and hence behaviors) are altered. In such an atmosphere morale is not always superficially sweet, but it is sound and real, and experienced as a significant anchorage.

We feel that in this attempt to apply client-centered principles to administration we have much to learn, but we feel that such application holds the exciting possibility of discovering basic scientific truth in regard to principles of interpersonal organizational relationships, truths which may greatly enrich and deepen our concept of democracy.

THE THERAPEUTIC HANDLING OF DISCIPLINE¹

ROBERT J. MINER

Director of Student Affairs, Miami University

EVER since man ceased being an isolated brute back in the early ages, and began to live with others in a more organized social structure, there have been constant deviations in his behavior from acceptable communal standards. These anti-social aberrations have met not only with disapproval from the group as a whole, but also with planned reprisal or punishment from leaders of the group in behalf of the group.

To prevent continued misbehavior, laws were gradually formulated, and methods for disciplining wrongdoers were devised. Today discipline is an accepted, if not a popular, social control in our scheme of complicated living.

What is discipline? The dictionary defines it as a system of rules controlling conduct; as corrective measures or punishment. All too frequently punishment is the only connotation aroused by the word discipline. To most people disciplining a person means punishing him. Our school system is no exception to this point of view. Even on the college level punitive discipline has been, and in many cases still is, the administrator's conception of how best to correct and prevent undesirable conduct. Lloyd-Jones and Smith point this out effectively:

The system of rules and penalties which now exists in most of our institutions probably constitutes a kind of hand-washing gesture on the part of the administration which either does not desire or does not know how to handle discipline by methods more truly educational. One readily concedes that it is relatively simple to write down detailed directions, which, if meticulously obeyed, would result in a minimum of social friction. Any office clerk can also make up a list of penalties for the infraction of each of these rules. It then becomes a relatively simple matter for any dean or faculty committee, no mat-

¹ Discussants whose comments are not recorded are Stephen M. Corey, Professor of Education, University of Chicago; Lincoln B. Hale, President, Evansville College; Margaret Holmes, student, Antioch College.

ter how lacking in imagination and teaching skill, to assign punishment to individuals in terms of rules and punishments, without much jeopardy to the punisher's feeling that his duty to society is being righteously fulfilled.²

"How can we eradicate objectionable behavior," one dean asked, "or put a stop to infractions of rules if we do not take a firm stand and let the students see by dire example the penalty they will suffer if they disregard standards of good taste or flaunt university regulations?"

Another dean replied, "Administer your discipline in such a way, that your offender ceases thinking that his real mistake was getting caught and realizes instead that his mistake was doing something undesirable. Discipline must help the student to understand why he acted as he did, and must assist him in deciding for himself whether this type of behavior befits him as an intelligent, maturing young man."

Lloyd-Jones and Smith elaborate this theory:

Anyone with experience in any of the better behavior clinics would understand that it may be possible to secure specific outward conformity by exerting external authority but that, if the behavior failure is due either to ignorance or lack of skill or to some underlying emotional cause, it is futile to expect results of a lasting character through the imposition of a group's authority. The chances are, in fact, that real harm is done to the individual by this course insofar as any genuine possibility for reeducation and rehabilitation are concerned.³

The modern concept of college discipline is predicated on the student personnel philosophy that discipline is an educative process, corrective, and not punitive. It should be a learning experience affording the student every opportunity for understanding himself better. It should also play a vital role in helping him adjust with greater facility to behavior patterns more acceptable than those which originally got him into trouble.

College students always have, and always will get into trouble. This is not abnormal, for in the main the problems which beset this particular age group are normal problems of adjustment, heightened and accentuated by the peculiar and

² Lloyd-Jones, Father and Smith, Margaret. *A Student Personnel Program for Higher Education*. New York: McGraw-Hill Book Company, 1930, p. 121.
Ibid., p. 126

unnatural environment which constitutes a university community. Such a community, especially in a small college town, comprises a concentrated group of young men and women, all within the same age limits, all undergoing the same process of intellectual, emotional, and physical maturation, and all temporarily living, working, eating, and sleeping in a restricted framework of adult imposed regulations.

These young people are still unsure of their capacities, yet eager to be considered independent and responsible. Their hopes run high, their ambition is unbounded, and their physical energy is unquenchable. They are highly competitive in their quest for status, tremendously uneasy in their search for long range security, and ever restless because their urge for mating is not fully sublimated or satisfied.

The oneness of this college group in such respects has, of course, been disrupted in recent years by the widening of the group's age span, and the introduction to classroom routine and campus mores of older men and women, many of whom are married, and most of whom have matured overnight in uniform in far-flung camps, or battlefields. This influx of veterans, with its consequent skewing of the campus age span, has in itself brought on new problems and multiplied old ones.

There always have been, and probably always will be certain areas of campus disturbances: Cheating in examinations, with perhaps such occasional tactics as ransacking offices and professors' homes for advance copies of final examinations, is a constant headache.

Thievery flourishes: The library steadily loses books; lockers are broken into and rifled; fraternity houses are systematically robbed; residence halls have their spasmodic waves of light-fingered pilfering; students "lift" souvenirs from stores in pranks which the courts call larceny; text books disappear when students leave them lying around unprotected.

Sexual misdemeanors periodically occur. It seems impossible to eliminate window peeping, genital exposure, pregnancies, homosexual and heterosexual irregularities, many of them adolescent, some of them pathological.

Drunkenness, gambling, destruction of property, and objectionable, disorderly conduct stalk along just often enough to

make everybody aware of their existence and their need for correction.

University regulations are constantly being abrogated. Residence hall rules are broken; student privileges are abused, "no parking" and "no smoking" requests are disregarded; grades on the registrar's official records are falsified. Some times such behavior is willful, sometimes accidental.

In the area of inter-personal relationships many incidents regularly happen which the head of residence, a dean, or some member of the faculty handle as a counseling situation rather than as a disciplinary act. These misdemeanors are not reported through the usual disciplinary channels because they do not seem to be serious enough to warrant standard disciplinary procedure, and generally they are wisely handled. Actually all discipline should be thought of as counseling, and should be administered as such, in order that it may be a therapeutic factor for the offender in his reconstruction of better attitudes and his establishment of more acceptable behavior patterns.

There are sharp differences of opinion about this concept. The disciplinarians in many colleges, whether they be individuals or committees, maintain that a university is a place of learning, not a corrective institution, and insist that reclamation of a weak or wayward youth is not the university's job. They declare that education is a privilege, and that the university is doing a real job of education for the many when it firmly shows how privileges may be lost to those who prove by their conduct that they do not deserve them.

Especially is this apt to be true of smaller institutions which are located in such little towns that everybody, townspeople and students alike, knows when John Smith spent the night in jail for being drunk. What kind of a boy he is, why he got drunk, and how he may behave thereafter apparently cannot be of as much concern to the university as what the students and townspeople may think if John isn't punished, and the campus made aware again that the "wages of sin are death." In maintaining cordial public relations in the town and respected standards on the campus, it is sometimes hard for the university to remember that it must also maintain its responsi-

bility to its individual students. Once a student has been accepted, the university assumes a definite obligation to help him, even when he makes mistakes, in his adjustment to his new environment, and in his quest for the kind of an education that will afford him maximum improvement in every area of his personal as well as his intellectual development. This obligation assumes that the welfare of the individual will not be sacrificed when he gets into trouble, for getting into trouble, and getting out of it are parts of the learning process which train young people how to live and work in greater harmony and peace with each other. Careful therapeutic discipline can help the individual effect this reorganization of his behavior attitudes without damaging the morale of his contemporaries.

Sociologists have made studies to ascertain the effectiveness of punishment in deterring delinquent or anti-social behavior. The findings of these studies tend to prove that whereas punishment may check petty and minor irregularities, it has little effect in eliminating seriously objectionable behavior. It does, however, have noticeable effect in making people cautious and more deft in the art of not getting caught.

Walter C. Reckless in his book *Criminal Behavior*⁴ states:

While punishment for deterrence appears to be logically efficacious, there is considerable doubt that it has as much deterrent value as the justification presupposes. Even in the era when extremely severe punishment was used for crimes of minor importance, no evidence can be found to support the view that punitive measures materially curtailed the volume of crime. . . .

For the great mass of infractions of law, the fear of consequences or the fear of punishment probably enters very little into causation. The conception of deterrence presumes that the person thinks before he acts, and that all he has to do is to think of the consequences; and then he will be deterred. Actually, however, individuals behave, act in certain ways, get involved in certain activities without the fear of punishment being held uppermost in mind.

Many institutions are uneasy and unhappy about their disciplinary procedures. They believe in the sanctity of the individual, yet they must protect the welfare of the whole

⁴ Reckless, Walter C. *Criminal Behavior*. New York: McGraw Hill Book Company, 1940, p. 273

student body. They recognize that an education includes emotional and social maturation, so they endeavor to furnish those curricular and extracurricular opportunities which make it possible for a student to learn how to live acceptably with his peers in conformity with those social and ethical restrictions which are necessarily essential to the well-being and survival of his group. They know the frailties of human beings, particularly of youngsters who are trying through experimentation, trial and error, to build self-confidence and a sense of personal security. They hope their educational program will assist their students in attaining this sense of adequacy by guiding them to a realization of their weaknesses, and an understanding of how to correct them. They know that therapeutic discipline is part of this process.

Why, then, do not more institutions administer this type of discipline? For one thing, they may not have enough money to hire a staff of psychologically trained experts as disciplinary counselors. They may not have a mental hygiene clinic to furnish psychiatric diagnosis and treatment. They hesitate to prescribe treatment for an individual if they have not a clinical psychologist who cannot only uncover the obscure and often deeply hidden causes of his trouble, but then work with him in correcting these traits. Or, in the last analysis, they may not know exactly how to go about exercising their disciplinary functions in an educative and therapeutic manner.

Whoever has this responsibility, and it may rest solely with the President, or be vested in the Dean, or a special committee, and whatever the practice may have been heretofore in administering it, discipline can be one aspect of the educative function, and it can become a constructive counseling experience for the student concerned without the expenditure of additional funds, or the acquisition of additional staff. A university is fortunate if it can afford to hire specially trained experts to handle its disciplinary problems. However, the many universities which cannot afford to do this can and should take account of their immediate resources, and should use their best qualified personnel in establishing a functionable, therapeutic disciplinary procedure.

First, and foremost, the President and the Faculty may need

to be indoctrinated and imbued with the educational concept of discipline. Understanding this, they may need to be encouraged to place the administering of discipline in the hands of a group of those faculty members and students who are best fitted by training, ability, willingness, and interest to devote their consecrated effort to a careful performance of this duty.

Having vested this authority where it can best function, the Faculty should then not take issue with the actions and decisions of this board. The board should have ample opportunity to try out its policies and procedures unhampered by advice or criticism.

It is desirable to have the Director of Student Affairs, or one of the Deans of Students serve as a non-voting member on this board. Usually this is the person who furnishes, in case-study form, most of the pertinent data concerning the character, personality, and background of the offender, and who comes before the board as the one best qualified to help that group thoroughly understand the student. After a decision is rendered, this is the person who helps the offender to comprehend and to accept the verdict. In fact, in some universities he has to handle every phase of discipline by himself without the help of any committee.

Not only must the Faculty be educated to think constructively about discipline, but the student body must at the same time have its attitudes toward discipline reshaped. If John Jones does not get kicked out for breaking a store window while he was drunk, the students find it difficult to understand why John Brown should be suspended for using his car without a permit. They, too, need to be trained to consider discipline as a learning process with penalties levied to fit the individual rather than the crime. They need to realize, and be glad of the fact, that a student in trouble gets every consideration from the disciplinary board, and they must understand that the board's decision is based upon its complete knowledge of the individual, and all the extenuating circumstances surrounding the case—knowledge which cannot be made available to others.

They need further to realize, if suspension is the verdict, that suspension is probably the best way the offending student can be helped to adjust more maturely to his responsibilities in the

world of which he is a somewhat confused member. They need to learn how to accept the action of the disciplinary board without gossip and condemnation. All this takes time—time, and careful, patient guidance.

The final step in this establishing of a functionable, therapeutic disciplinary procedure is the actual procedure used for the handling of each case. This should be so designed that the individual is afforded a maximum amount of guidance and protection, and it should be standard practice.

First of all, the student concerned should be interviewed to obtain from him his statement of what occurred and why. The accuser, or whoever reported the misbehavior, should likewise be consulted for clear and accurate information about the specific nature of the particular offense. Occasionally it is possible that the accuser has acted in rancor, or for spite, or for personal reasons involving prejudice, or malice, or gross misunderstanding. All of this must be known, everything possible about the case must be thoroughly comprehended, for sometimes the student must be protected as well as helped.

The agent who works with the student in this sort of an organization is actually a counselor, and he functions better if he has had specific psychological training in the techniques and skills of interviewing, diagnosing, counseling, and therapy. It is his job to explore with the offender the nature of his offense, and the personality traits, and emotional factors which contributed to it. He must help the student understand why he behaved as he did, and at the same time he must create for the offender such an atmosphere of acceptance and security that the student will be encouraged to want to function more satisfactorily within the established social framework of which he is a part.

This counseling process is of itself therapeutic, for it helps to maintain the self-respect of the student. When a person is in trouble he needs desperately to be certain that he is not going to be rejected by his peers because of his misconduct. His acceptance by his counselor is reassuring, and gives him the courage to face himself, his troubles, and his associates. As he talks with his counselor, and brings to light the many conflicts which may have beset him, perhaps since early childhood,

he begins to see how various pressures have built up, and how his behavior has been conditioned by his repeated reactions to these forces. With such insight comes an appreciation of the dynamics of his present conduct, and generally a determination to want to change it to ways more acceptable to his fellows and to himself.

This is not an instant process. The counselor utilizes every available source to accumulate helpful information about the student. Intelligence and achievement test scores are studied. The student's scholastic performance is examined in the light of these scores. Oftentimes the tremendous difference between performance and ability indicates the presence of anxieties worth investigating.

The Head of Residence, the Faculty Adviser, the student's various instructors, and his Fraternity Adviser are consulted. If individual personnel records are kept these are carefully explored for personal history which may add to the counselor's understanding of the ability, personality, and developmental background of the student. Repeated interviews with the student take place interviews which are continued long enough after insights have been obtained to make sure that the individual is securely started in his process of rehabilitation.

Every available resource is tapped in this process. The Health Service can indicate whether health hazards have contributed to the student's difficulties, and if so, how they can be improved or eliminated. The mental hygiene clinic can diagnose whether the individual may be neurotic or psychopathic, and can furnish specialized psychiatric treatment if it is indicated by the diagnosis. Teachers can be encouraged, if necessary, to use different classroom tactics with the individual. Opportunities for developing greater social competence and a corresponding sense of responsibility can be provided.

If discipline is really educative and follows a therapeutic procedure, the offender will not be scolded, ridiculed, or denounced for his misdemeanor, no matter what the offense may be. If the misconduct is such that actual penalties have to be invoked, care should be taken to make sure the student comprehends why the penalties have been prescribed, and why

he must submit to them. If suspension is the only verdict available the student must be helped to understand why the university cannot condone his conduct, why the existing disciplinary set-up does not permit any different handling of the case, and why the university regrets having to dismiss him. Penalties will not be inflicted merely as standard procedure but when they are invoked they will be chosen as a definite part of the therapy. They will not be a punishment for a certain crime, but a benefit for a certain individual. These steps in themselves are constructive, and are an investment in future sound public relations as far as the university is concerned. As far as the individual is concerned, such treatment leads to a better emotional adjustment, and the preservation of self-respect.

Cooley has stated that the community is extremely unwise to take any action that will destroy the self-respect of the offender,⁵ and Van Waters asserts that the self-respect of the offender is the basis of all successful efforts toward his rehabilitation.⁶

After the disciplinary board has finished with its hearing and its verdict, the student cannot be left alone to flounder. It is preferable to have the one who has worked with him so far as a counselor continue in that capacity, but if this is impossible the student should be referred to some sympathetic, sensitive, competent staff member who is willing to serve as a friend and guide, whether the student stays in school or leaves on suspension. If the offending student needs a type of specialized therapeutic service which is not obtainable at the university, he should be referred to proper agencies nearest his home, and be removed from the university until improvement in his condition is indicated.

To summarize what has thus far been stated: Many collegiate institutions still prefer the old concept of punitive discipline to the newer philosophy of therapeutic correction. Some of those institutions which cleave to the punitive attitude do so because they feel the penalties they inflict on the individual tend to safeguard the entire student body as a whole,

⁵ Cooley, C. H. *Life and the Student*. New York: A. A. Knopf, 1927, p. 204.

⁶ Van Waters, Miriam. "The Delinquent Attitude." *The Family*, V (1924), 112

and keep the majority more completely under control. Others do so because this method involves less work than the case study, therapeutic approach. Some institutions excuse their lack of corrective procedure on the grounds that a staff of experts is a financial impossibility.

However, an increasing number of universities are administering their discipline as a function of their educative process. If funds are available, their organization is staffed with psychologically trained specialists. If funds are scarce, the best of their regular personnel make certain that all offenders get a chance at rehabilitation under some sort of guidance.

If a university believes in the therapeutic disciplining of its misbehaving students, it will find a way to help them profit from their mistakes, and adjust to better behavior habits. If a university does not recognize any such responsibility toward its offenders, then the guilty ones can expect to be punished in a manner befitting the trouble they have caused.

Four freshmen boys set forth one Saturday night in quest of fun and adventure. They didn't have dates for their dormitory dance, so they decided to go to a neighboring town, and explore some of the night spots. They knew that using Ned's car for such a jaunt was contrary to university regulations, but who would ever find out about it?

Late Sunday afternoon one of these boys phoned long distance from the police station where they had all spent the night, and humbly asked the Dean of Students if he could possibly get them out of jail. The Dean went to that town, conferred with the Chief of Police, signed for the four boys, and brought them home. Two days later he went to court, and stood before the judge with them while they were tried on charges of disorderly conduct and grand larceny. They had appropriated as souvenirs some expensive bar equipment from one of the "joints" they had visited, and all of them had consumed too much beer and whiskey.

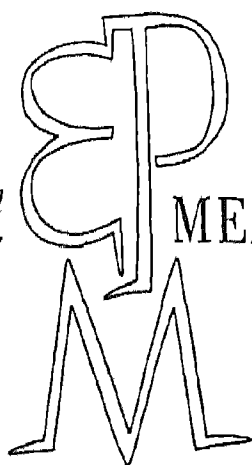
When the judge dismissed all charges against them and put them on probation for one year to the Dean of Students, he explained that whereas he did not condone their conduct one bit, he did recognize that they were immature youngsters whose poor judgment had landed them in the hands of the law. He didn't want them to have a court conviction, for he was sure their experience would keep them from becoming repeaters.

The judge was correct. Their stay in jail had been disturbing and sobering. All of them had regretted exceedingly their indiscretion, and each of them had determined never to get into any mess again. They were ashamed and humiliated.

When they appeared before their university disciplinary committee, their heads of residence, their advisers, the Dean spoke in their behalf. They had been excellent campus residents, and they were all fine lads. A plea was made that they be allowed to continue at the university on probation to the Dean or some other reliable staff member. They were all suspended for the balance of the semester, the faculty members of the committee outvoting the student members, who held out for probation. The decision was purely punitive, a far cry from anything constructive or therapeutic.

Unfortunately, when students come before disciplinary committees to make restitution for the commission of sins, such verdicts still happen too often, and in too many institutions. "Madame LeFarge" again sits in session, and there is *knitting at the guillotine*.

EDUCATIONAL and
PSYCHOLOGICAL



MEASUREMENT

VOLUME EIGHT, NUMBER FOUR, WINTER, 1948

<i>Questioning Some Assumptions Underlying Current Achievement Testing.</i> VERNER M. SIMS	565
<i>Selection of Physical Scientists.</i> MILTON M. MANDELL and SIDNEY ADAMS	575
<i>A Factor Analysis of Selected Senate Roll Calls, 80th Congress.</i> CHESTER W. HARRIS	583
<i>Some Observations on the College Board Language Tests.</i> HENRY S. DYER	593
<i>The Group Origins of Student Leaders.</i> E. G. WILLIAMSON	603
<i>The Dependence of Factorial Composition of Aptitude Tests upon Population Differences among Pilot Trainees. I. The Isolation of Factors.</i> FRANK J. DUDEK	613
<i>The Role of Client Talk in the Counseling Interview.</i> EARL F. CARNES and FRANCIS P. ROBINSON	635
<i>The Relationship Between Values, Concepts, and Attitudes.</i> ASAHIEL D. WOODRUFF and FRANCIS J. DIVESTA	645
<i>Measurement in Rural Housing—A Progress Report.</i> CHARLES L. DURRANCE, JR.	661
<i>Intelligence Test Results Obtained from a Specific Type of Army A.W.O.L.</i> JERRY H. CLARK	677
<i>Predictive Values of the Stanford Scientific and the Engineering and Physical Science Aptitude Tests.</i> H. A. COOPRIDER and H. R. LASLETT	683
<i>The Contributors</i>	689

ERRATUM

The first two sentences of the second paragraph, page 174, of C. Harold Stone's article, "Are Vocational Orientation Courses Worth their Salt?" which appeared in the Summer (1948) issue of this journal, should be revised to read. "For counseled students, trends in the laboratory groups (Experimental II and III) were in the same direction as for the lecture group (Experimental I). However, fewer laboratory students who were not counseled had poor choices at the end of the year and there was an increase in optimal choices during the year."

QUESTIONING SOME ASSUMPTIONS UNDERLYING CURRENT ACHIEVEMENT TESTING

VERNER M. SIMS
University of Alabama

THE value of any measurement of educational achievement must be judged ultimately in terms of the extent to which the measurement reveals information leading to a more intelligent prediction or control of the behavior of students. Commonly, however, we neither apply this criterion directly to our instruments nor to our measuring activity. Instead, in the testing which we do, we either consciously or unconsciously accept a chain of assumptions which stem from and presumably may be justified, either logically or experimentally, as supporting this ultimate purpose. The soundness of our measurement and its usefulness to education is, therefore, contingent on the validity of a set of assumptions which are implicit in our testing. It is the purpose of this paper to make explicit and to critically examine certain of the assumptions which seem to underlie current measurement of achievement.

1. Current measurement of achievement assumes that the value of a learning experience is indicated by increased ability (skill or knowledge) to cope with some situation or class of situations. <

Achievement testing as presently practiced appears to be confined to the testing of ability. Standardized tests of achievement, whether subject tests or general tests, survey or diagnostic, whether tests of knowledge, skills, or understanding, or tests of ability to apply principles and interpret data, are all concerned with measuring what children *can* do. Teachers in their own testing and marking are concerned with measuring the same outcomes. If the teacher does attempt measurement of anything other than ability it is usually made clear to the pupils that "this test does not count." For teachers to permit anything other than achievement (interpreted as in-

creased knowledge or ability) to influence their marks is commonly considered questionable practice.

- The obvious conclusion to be drawn from an examination of current achievement testing is that the end of education is to increase the abilities of pupils. Most of our educational theorists would not be willing, however, to accept any such conclusion. They would say, rather, that the end sought in teaching is changed behavior on the part of the learner; which is, of course, not the same thing as having the ability to do differently. Whether one *will* act so and so depends not only on whether he *can* do so but also on a complex of attitudinal-emotional-motivational factors which are generally neglected. Furthermore, it seems reasonable to assume that whether the learner will use his learning or not is a factor of the setting in which it is acquired. Not only the learner's own purposes and methods but those of his teacher will condition the kind and amount of use made of the learning.¹

If the fundamental question in achievement testing is whether, in terms of the evidence at hand, one can predict that the learner will act differently in a certain class of situations, it would seem desirable that our testing comprehend more than the measurement of *ability*. Recognition of this assumption and acceptance of its limiting influence would, in the writer's opinion, lead to a re-vamping of many of our present approaches, to the designing of new instruments, and to a re-interpretation of many of our test findings. Tests might be designed, for example, to measure whether a child *would*, as well as could, read, reason or act in certain types of situations. Measures of the effectiveness of the use made of learnings might be developed. In time we might be able to describe (even quantitatively) the conditions under which learnings would or would not function.

- 2. Current measurement of achievement assumes that the value of a learning experience is indicated by growth in those directions toward which the learning (or teaching) activities are specifically pointed.

¹ The writer has developed this point at considerable length in an earlier article, "Educational Measurements and Evaluation," *Journal of Educational Research*, September, 1944, pp. 18-24.

It seems fairly obvious that this assumption does underlie our testing. Typically, we test the value of drill in spelling by measuring the increase in ability to spell, the value of instruction in health by measuring the increase in knowledge and understandings relating to matters of healthful living, the value of a method of science instruction by measuring the increased ability to apply scientific principles, etc. We measure the child's growth and mark him on the achievement of the specific objectives which are set up for the course.

Superficially the soundness of this procedure would seem almost axiomatic; but the thoughtful reader will recognize immediately that it is justified only when many other things are accepted as "being equal," and without evidence to support this, one cannot always assume that such is the truth. If the increased ability to spell comes as a result of sacrifice in the ability to read, if it is accompanied by the acquisition of unfavorable attitudes toward spelling or toward school, if it furnishes an escape from real problems about which the children could and should be doing something directly, or if it merely results in pupils becoming increasingly better in a skill in which they are already good, one would certainly question the value of the learning experience. Yet evidence relating to such matters is commonly not made available. Such information is usually not even considered the concern of the tester.

If the "whole child" is involved in any learning experience, then inferences concerning the value of learning which ignore this fact are of questionable validity. Testing theory which neglects the fact forces the trained user to interpret his test results in conjunction with a large amount of "common sense," subjective and informal data, and encourages the untrained user to arrive at conclusions that are unsound.

3. Current measurement of achievement assumes that the value of a learning experience may be inferred from measures of the outcomes evident at some particular moment after learning, typically at the conclusion of a learning experience. <

The test-teach-test formula describes the accepted procedure for evaluating the outcomes of learning. In practical classroom situations and in educational experimentation we test the value of method and the growth of children in such terms. Progress

is assumed to be measured by noting the difference between initial and terminal test scores. Furthermore, in cases where there has been no previous formal training the first testing is commonly dispensed with, and the terminal test score becomes the measure of learning.

We do this although we know that learning is a form of growth the path of which is best expressed by a curve. If this is true, predicting later performance is a matter of projecting this curve into the future. We attempt to do this extrapolation by locating one point (in the case of terminal testing only) or at the best two points (in the case of initial-terminal testing) on the curve. Mathematically, this is not possible. With one point we can tell nothing about the curve, with two points the curve would need to be a smooth one the slope of which were known. These are conditions which are not met in practical teaching-testing situations, unless in the doubtful case where the concern is entirely with group averages. Certainly no individual's curve of learning is a "smooth one the slope of which is known." This implies that dependable predictions can be made only when we locate further points on the curve of growth.

Some educators will point out, however, that there is a correlation between such measures of progress and later performance. The fact is, the nature of these correlations is one of the reasons which cause the writer to question the soundness of this assumption. The size of the coefficients usually found between terminal tests and later achievement suggest that we may be dealing with little more than "the correlation between desirable traits." Reading age, at ten years, for example, is no more closely correlated with reading ability at 15 years than is general intelligence. In the writer's institution at least, terminal achievement on a standardized test of high-school algebra is not as closely correlated with success in college algebra as is general quantitative ability as measured by the *ACE Psychological Examination*. One is led to suspect that the only real justification for assuming that initial-terminal testing measures the value of a learning experience is found in the convenience and simplicity of the procedures which the assumption makes possible.

4. Current measurement of achievement assumes that the only sound interpretation of measures of achievement is through comparison with norms of some sort.

Educational measurement, as it has developed, is a "normative" science; that is, it attempts to develop norms of performance for groups of known (and described) status from which one may by comparison interpret measures of other groups or individuals. Standardized achievement testing is universally of this nature. Raw scores are to be interpreted in terms of age or grade norms, T-scores, scaled scores, or percentiles. The writer knows of no standardized achievement test which does not use this method of interpreting scores. In fact, the word "standardized" has come to refer not only to a standard testing procedure but to a test with "standards." Furthermore, very few tests propose any other method of interpretation. In teacher testing, too, the emphasis, as the testing experts would have it certainly, is in the same direction. In all text books concerned with the problem of marking with which the writer is familiar, teachers are encouraged to make relative interpretations, to "mark on the curve," to define an A as superior performance in the group; some even go so far as to recommend that promotion and non-promotion be determined on the basis of position in some group. Any teacher who proceeds otherwise today is apt to be labeled either old-fashioned or too "progressive."

The writer is not questioning the value of the interpretation of test scores in terms of norms. In the hands of a trained teacher, standard scores may serve many useful purposes. (The fact that they are too often misused, being accepted as goals of achievement, for instance, is also irrelevant to our purpose here.) Relative position in a class, as expressed by the kind of marks which specialists in measurement advocate, also has its uses. That which is being questioned is the assumption that the *only* valid interpretation of measures of achievement is through comparison with norms. For many school objectives and for many situations facing teachers, the writer contends that there are other perfectly valid interpretations possible and desirable. There are two cases in particular which seem worthy of comment.

First, in the case of some learnings, worthwhile interpretations may be made by comparing a child's performance with the demands of life (present or future). For example, we assume that a child needs to know how to add, and to know automati-

cally, all the digit combinations. The fact that Johnny has mastered 50 per cent of these combinations, or 60 per cent, or 90 per cent, is valuable information for the teacher who is working with Johnny without regard to where he stands with reference to other children. In fact, the job which lies ahead for Johnny is just the same whether he is at the top, in the middle, or at the bottom of some group. Or, let us consider writing or spelling. Johnny must write legibly (we assume), and whether his writing is legible, or not, can be determined without comparing him with other children. If we assume that the child should be able to spell the 1,000 most common words, or
 > the 100 "spelling demons," then knowing that he can spell certain of these words has great meaning to Johnny and to his teacher, even though he is the only pupil she has ever had in class or even known. If there is anything wrong with this idea it is in the assumption concerning what he should know, not in the interpretation of the measurement.

Our whole system of prerequisites falls into the same category. If certain learnings are really needed to succeed in later work then tests which show whether the student has mastered these learnings have meaning which is independent of the performance of others. The same is true of learning in multitudinous fields. Particularly in the fields of attitudes and habits of action could illustrations be multiplied. The fact that a child throws rocks at Negro children every time he gets a chance, the regularity with which he brushes his teeth, or the extent to which he eats a balanced diet, all have meaning which is not determined by a normative score.

Secondly, many measurements may have meaning when interpreted in terms of the student's own educational-psychological make-up. Illustrated simply, the child who succeeds in mastering the mechanics of arithmetic but who cannot solve problems involving the same processes can be located without any normative interpretations. The student whose factual learning is entirely acceptable to the teacher but whose attitudes toward the same problems are not; the student whose emotional blockings prevent him from using his ability to reason sensibly on problems relating to labor unions; the student whose motivations are all in the direction of pleasing the teacher:

These can all be identified without the use of norms, whether class norms or large group ones. The fact that the teacher's own experiences with students are involved in the judgments which she makes concerning the goodness or badness of such characteristics is aside from the point. Exactly the same value judgments are involved in the interpretation of tests which are provided with norms.

Insisting that measures of achievement have meaning only through comparison with norms must result in limiting the usefulness of educational measurements.

5. Current measurement of achievement assumes either that "appropriate" measurement is independent of one's theory of learning and of education, that current measurement concepts fit all current theories of learning and of education, or that there is only one acceptable theory of learning and of education. L

When achievement testing techniques were first subjected to careful study, some three decades ago, the prevailing theory of learning was a narrow, "connectionist" one, expressed by Thorndike's stimulus-response formula, and the commonly accepted theory of education was that education is preparation for living. It is perhaps no accident, therefore, that our testing procedures so aptly fit these theories of learning and of education. In the meantime, however, in many quarters at least, other theories of learning and of education have come to be accepted, while our testing techniques remain relatively unchanged. (The only major change in achievement testing procedures that has taken place during the past two decades which the writer is able to identify is represented by Tyler's approach. Most of our testing seems to fit the theory that education is an accumulation of large numbers of relatively independent knowledges and skills, while Tyler's testing would seem to identify it as an accumulation of a large number of abilities to apply principles, interpret data, and reason logically, all of which are still relatively independent. It is not difficult to see in Tyler's work the influence of Judd's theory of "transfer through generalization" rather than Thorndike's "transfer of identical elements," but otherwise it fits perfectly an atomistic conception of learning as preparation.)

So far as the writer can find, no one working in the field of


achievement testing has seriously considered the implications of insightful learning, or of organismic or "field" psychology for achievement testing procedures. Nor does anyone seem to have considered the possibility that other theories of education, such as the theory that education is a form of intelligent living which within itself constitutes an end, may demand different approaches to the matter of testing.

If one rejects the theory that learning consists of the acquisition of a large number of more or less simple and independent abilities, then one can justify present means of testing the outcomes of instruction only by assuming that the number of such abilities acquired is indicative of total learning. When the fundamental nature of the differences among the several theories of learning and of education is considered it seems rather naive to accept such an assumption without any evidence to support it, even without trying to get evidence to support it. Can we assume, for example, that a child who showed no progress on a current standardized reading test during his first three years in school, and then, in his fourth year, without any special help, jumped from a reading age of seven years to almost eleven years, actually learned nothing about reading during the first three years?² How could one who denies that psychologically the whole is equal to the sum of the parts be expected to take to current tests of arithmetic, of social studies, of science? As long as we operate as though the end of education is the accumulation of a set of skills and knowledges which will be useful later, can we blame the "progressives" (who accept a quite different theory of education) if they reject our instruments in their entirety?

It does not seem unfair to say that the measurements experts have very subtly, although perhaps unconsciously, thrown their support to a particular theory of learning and of education. Would not an eclectic approach contribute to a greater serviceableness from the discipline? With our present lack of effort, can we presume to say that educational measurements have no contribution to make to those who hold to differing philosophies of education or belong to differing schools of psychology?

² Reported by Willard C. Olson, "When Should My Child Learn to Read?", *University of Michigan School of Education Bulletin*, XIX(1947), 9.

In summary, the attempt has been made in this paper to point out several assumptions which seem to be implicit in present-day measurement of achievement. The assumptions identified are those of which the writer is critical. In general, the basis of the criticism is the fact that through operating on the assumptions we restrict the usefulness of educational measurements. Whether or not the assumptions are defensible may, of course, be debated. If they do underlie our measurements, however, there can surely be no question concerning the worth of making them explicit, and the desirability of consciously recognizing them as assumptions. It may even be that simply accepting the assumptions for what they are would, in itself, result in some rather fundamental changes in our measuring activity and in the techniques of measurement finally developed.



SELECTION OF PHYSICAL SCIENTISTS¹

MILTON M. MANDELL and SIDNEY ADAMS

U. S. Civil Service Commission

THE development of valid methods for the selection of scientists for industry and government is a *must* project for those engaged in personnel selection. All nations have learned that the status of scientific inquiry is an important ingredient of both national defense and prosperity. All of our scientists, military, industrial, and university, contribute to these important goals.

Little evidence of a statistical nature exists regarding the validity of methods currently used for the selection of physical scientists for employment. Universities have studied the question of predicting scholastic achievement in the physical sciences. There is little or no evidence, however, to indicate that the tests used for this purpose will predict employment success.

The United States Civil Service Commission began on November 1, 1947, a study of selection methods for physical, biological, and agricultural scientists. The President's Scientific Research Board, in its report of October 4, 1947 on *Administration for Research*, had recommended that "The Interdepartmental Committee on Scientific Research and Development should sponsor a series of studies to be carried out by departmental personnel staff and Civil Service Commission staff relating to the broad problems of . . . examining and testing techniques for scientific positions . . ." (p. 171). Following is a preliminary report on the work done during the first six months of this project. This introductory statement has been prepared now in the hope that the information contained in it will help to stimulate additional work by others in this field. This report is limited to a discussion of selection methods for physi-

¹ "Physical scientists" for the purpose of this article include physicists, chemists, and research engineers.

cists, chemists, and research engineers; it does not include any discussion of biological and agricultural scientists.

Job Analysis

Information regarding the duties of physical scientists was obtained by the use of the following methods: (1) Discussion with leading scientists in the Federal Government; (2) review of the available literature; (3) review of classification specifications for physical science positions; and (4) preparation by a small group of physical scientists of written statements describing their duties. The following tentative test battery was based on the information obtained from these four sources.

Test Battery

The following tests were used in the attempt to measure the abilities indicated as essential by job analysis. No profound analysis is needed to determine that other skills also need to be measured, in addition to those presumably measured by these tests.

1. *Abstract Reasoning (Figure Analogies).*—This five-choice test attempts to measure abstract reasoning ability. It has been widely used by the United States Civil Service Commission for entrance level professional and administrative positions; the results obtained have been consistently good.

2. *Gottschaldt Figures.*—This is a revision by Professor L. L. Thurstone of items originally developed in Germany. The factorial composition of this test, which requires the detection of hidden geometric figures, has not been fully determined.

3. *Spatial Tests.*—The following types of spatial items are now under study: (a) Cube turning; (b) surface development; (c) two-cube (two separate cubes with various designs on their faces are presented; one cube has two prongs extending from one face—the other cube has two holes on one face; the task is to determine which one of a number of sets of two cubes joined together would be formed by the two separate cubes if the cubes were joined by placing the prongs of one cube in the holes of the other); (d) block formulation (the task is to determine which one of a number of large blocks would be formed

by the combination of a specified number of three sizes of blocks)²; and (4) block counting.

4. *Mathematical Formulation*.—The task in this test is to express a narrative statement in terms of an equivalent mathematical formula

5. *Letter and Number Series*.—This test was included with the thought that the reasoning ability measured by the test might be important.

6. *Table Reading*.—This test was developed by the Air Forces. It was administered in this study as a speed test (7 minutes).

7. *Verbal Analogies*.—The inclusion of this test, which attempts to measure verbal and reasoning abilities, in the battery was suggested by Dr. M. H. Trytten of the National Research Council, who is Chairman of the Civil Service Commission's Advisory Committee on Scientific Personnel.

8. *Interpretation of Data*.—This test, devised by the Progressive Education Association, measures the ability to evaluate the accuracy of conclusions from statistical data. This test was scored by two methods: (a) The number of items right; and (b) the direction of the errors, that is, whether the errors tended to represent cautious or bold evaluation of the statements.

9. *Hypotheses*.—The task in this test, devised by Dr. Max Engelhart of the Chicago City Colleges, is the evaluation of the relation of facts to hypotheses, and the relation of hypotheses to the explanation of physical phenomena. This test has a high degree of face validity in selection work in the physical sciences.

10. *Scrambled Sentences*.—The task in this test is the determination of what the final word in a sentence would be if the words, which have been disarranged, were put back in their original order. Sentences about 15 words in length relating to the physical science field were used in this study.

11. *Biographical Information Blank*.—There is general agreement that interest in and enthusiasm for research work are

² This test was developed by B. J. Winer, formerly of the Test Development Unit staff at the Civil Service Commission.

essential ingredients of a successful career in the physical sciences (and other occupations also). The items in the BIB attempt to measure these traits as well as schooling and personal relations. Most of the items had originally been developed by the Department of the Army for military selection.

12. *Reading Comprehension.*—These are paragraph-reading items from the general science field. Proper interpretation of reports of scientific studies would seem to be an important skill to measure in this study.

14. *Science Judgment.*—This test is still in the process of development. It attempts to measure judgment in terms of such problems of science methodology as the design of experiments, and the interpretation of scientific findings. Its contents cover the field of the physical sciences; it does not attempt to be specific in terms of one or another subject.

15. *Subject-Matter.*—Two subject-matter tests have been prepared, one in the broad field of physics and one in chemistry. Because of the lack of staff to prepare a test in engineering, the physics test has been used as the subject-matter test for this group.

16. *Vocabulary.*—The items are general rather than specific to the sciences.

Experimental Groups

All or most of the tests described above have been administered to the following groups: (a) 37 first-, second-, and third-year graduate students in physics and chemistry at Catholic University; (b) about 250 physicists, chemists, and engineers, in grades P-1 through P-7 (about \$2,900 to \$9,600), engaged in testing, instrumentation, standardization, and other research work at the National Bureau of Standards and the Naval Ordnance Laboratory; and (c) about 125 research and development engineers at the Air Materiel Command laboratories at Wright-Patterson Fields.

Criterion

The criterion used at Catholic University was a rank-order listing by professors on over-all ability, theoretical ability, and experimental ability. The professors were asked to rate the

students on the basis of predicted performance as employees in a laboratory, rather than on University performance.

The criterion used at the Federal Government research laboratories was a graphic-rating scale consisting of a number of elements to be rated on a five-point scale, and summary ratings also made on a five-point scale, of general over-all ability, theoretical ability, and experimental ability. The ratings, specially prepared for the purposes of the study, were completed by supervisors, colleagues, and subordinates in two of the research laboratories; in the third laboratory, the subordinates did not rate. The ratings from all groups were combined arithmetically into one average rating for use as the criterion. Salary level was used as an additional criterion. Preliminary analysis of the agreement among raters at one laboratory, where the average number of raters per subject was low, indicates that the extent of agreement was somewhat higher for Chemists than for Physicists but the reliability was too low, even for Chemists, to expect satisfactory results.

Following is a brief statement regarding the value of these tests for predicting ability in the physical sciences. It should be recognized that there were too few items in many of these tests to obtain satisfactory reliability; also, the method of test administration was such that subjects could spend different amounts of time on each test since a number of tests were administered with one over-all time limit.

1 *Abstract Reasoning*.—The validity coefficients for this test for samples of between 21 and 28 Chemists and Engineers are about .5^a, using ratings as the criterion; the validity coefficients for Physicists, using ratings and salary level as the criteria, are about zero. Successful discrimination for this test was also obtained on the Catholic University student sample. It is difficult to reason why these different results for Physicists were obtained; the standard deviation of the test for a group of 38 Physicists was 9.04, which indicates that lack of spread in the scores would not account for this result.

2. *Gottschaldt Figures*.—Engineers were the only group for whom this test produced even moderately good results; using

^a All of the coefficients are product-moment.

ratings as the criterion, a validity coefficient of .4 was obtained for 28 Electrical Engineers.

3. *Spatial Tests*.—Surface development items gave moderately good results for Physicists and Chemists although these results are highly tentative; for Engineers, however, the block formulation items had a correlation with ratings of two groups, 26 Electrical Engineers and 19 Mechanical Engineers, of .4. None of the other spatial tests produced good results with these groups.

4. *Mathematical Formulation*. This test was one of the best tests among those tried out on Engineers; the validity coefficient for a group of 26 Electrical Engineers, using ratings as the criterion, was .59. The results with Chemists were moderately satisfactory; the test did not discriminate among the Physicists.

5. *Interpretation of Data*.—Promising results were obtained for this test with Chemists and Electrical Engineers; the results with other groups were insignificant.

6. *Hypotheses*.—This test was the only test whose results were promising with all of the three groups, Physicists, Chemists, and Engineers, included in this study; the high inter-correlation of this test with most of the tests, which probably accounts for its consistent validity, also tends to reduce its value since the zero order validity is not high enough to select on the basis of this one test alone. It may be possible to purify this test to reduce these intercorrelations without too much loss in validity.

7. *Biographical Information Blank*.—The results obtained from this instrument are very promising. The validity coefficients obtained for this Blank with Chemists and Physicists, using a pre-determined scoring key with weights of +1, 0, and -1, were moderately good; in addition, the low inter-correlation of the scores on this Blank with the other tests indicate that it can contribute to a multiple correlation. Based on a total of 98 physical scientists, consisting of two groups of Physicists and one group of Chemists, the following types of items seemed most promising: outstanding work in mathematics in high school; graduation from college in the top 10 per cent; lack of participation in high school athletics; graduation at the age of 16 from high school; and lack of interest in the administrative

aspects of scientific work. Items on these subjects, plus a number of others, were highly discriminating with the item of age of graduation from high school the most discriminating.

8. *Science Judgment*.—Based on performance ratings as a criterion, this test gave the best results on a sample of 33 Physicists, $+.48$; its results for Chemists were also moderately good, while the validity coefficients with Engineers were about zero. Since the Engineers tend to be engaged in development work, rather than research work, this difference can be understood. It is hoped that a more reliable test of this type would be useful for selection for research positions.

9. *Subject Matter*.—This test gave the most consistent results among the two independent samples of Physicists consisting of 62 cases at the Bureau of Standards and the Naval Ordnance Laboratory, using salary level as the criterion. The results in both cases were about $+.5$.

The other tests in this battery, based on the obtained data, did not contribute significantly. These data cannot be considered conclusive because the various factors that contribute to the results were not controlled sufficiently in these pilot studies. Studies based on these tentative results are now being continued so that firm data, useful for prediction of ability in these fields, may be obtained. The tentative results obtained seem to indicate that selection for this important group of positions can be improved.

A FACTOR ANALYSIS OF SELECTED SENATE ROLL CALLS, 80TH CONGRESS

CHESTER W. HARRIS

University of Wisconsin

IN SEPTEMBER, 1947, the League of Women Voters of the United States published a pamphlet entitled *Significant Roll Calls, 80th Congress, First Session*, in which they presented voting records on selected issues. In a foreword, they point out:

Since it is not possible for us to list the full record of every member of Congress on all issues, we have selected what we consider to be some of the most significant roll calls during the first session of the 80th Congress. The selection was made on the basis of issues in which the League has been active, with the exception of a few issues of general interest on which the organization took no position.¹

Using these data, an analysis has been made of the voting record of the Senate on nine issues. The procedure was to determine tetrachoric coefficients of correlation between each pair of roll calls for the 95 senators.² These correlations were then analyzed by means of factor analysis.

The Senate roll calls listed by the League of Women Voters are as follows:

1. Taft motion to table second Overton motion to seat Bilbo. Passed.
2. Tydings amendment to continue investigation of the national defense program under a standing committee (Executive Expenditure) rather than under a special committee. Defeated.
3. Bricker motion to recommit nomination of Lilienthal as chairman of the Atomic Energy Commission. Defeated.
4. Greek-Turkish aid bill. Passed.
5. Taft-Hartley labor bill. Passed.

¹ League of Women Voters of the United States, *Significant Roll Calls, 80th Congress, First Session*, 726 Jackson Place, N.W., Washington, D. C., September, 1947. (Publication No. 98, page 2.)

² Senator Bilbo (Mississippi) did not vote. Overton's motion to seat him was tabled on January 3, 1947. Senator Bilbo left Washington without being seated, and later that year, died.

6. Kem amendment to cut foreign relief funds from \$350 million to \$200 million. Defeated
7. Income tax reduction bill. Passed.³
8. Hawkes amendment to allow 15% increase in rent when agreed to by landlord and tenant. Passed.
9. Conference report on wool bill, providing price support and authorizing increase up to 50% on imported wool. Passed.
10. Resolution to disapprove the President's Reorganization Plan #3, consolidating the housing functions of the government in a single agency. Defeated.

Roll call number 1, above, was not included in this analysis, since a large proportion of the Senators had not been seated—and consequently were not eligible to vote when Taft's motion was introduced. The remaining nine roll calls were included, and in addition the party designation (Republican or

TABLE 1
Intercorrelations (Tetrachoric) of Nine Senate Voting Records and Party Designation
(N = 2295)

	P	2	3	4	5	6	7	8	9
P									
2	.99								
3	.72	.77							
4	.33	.46	.66						
5	.78	.82	.72	.62					
6	.15	.32	.75	.72	.41				
7	.94	.97	.80	.33	.88	.25			
8	.93	.97	.77	.20	.86	.15	.91		
9	.61	.64	.59	.30	.10	.24	.51	.40	
10	.82	.85	.79	.54	.52	.55	.79	.82	.73

Democrat) was included as a variable. This gave ten variables for analysis. The symbol *P* is used to designate the party variable. Numbers corresponding to those listed above are used to identify the nine roll calls. In determining the correlations, a formally-expressed opinion was regarded as equivalent to a vote. Congressional procedures provide for pairing for or against issues, and, in the Senate, for the announcing of a position for or against an issue on behalf of an absent Senator. Pairing and announcing thus give formal expression to a Senator's attitude toward or judgment of an issue.

The correlation matrix is presented in Table 1. In prepar-

³ This refers to the original passage of the Bill. Later it was vetoed and the veto sustained.

ing the correlations, a vote or opinion in favor of the issue was taken as the positive section of the dichotomy, and the designation of Republican was made the positive section of the Republican-Democrat dichotomy. When this is done, variables 2 and 6 have consistently negative relationships with each of the other variables. These two variables therefore were reflected to secure the matrix presented in Table 1. The interpretation of these correlations may be facilitated by listing the variables and the designation or opinion that represents the *positive* section of the dichotomy.

1. Republican.
2. Against Tydings amendment (i.e., in favor of setting up a special committee to continue the investigation of national defense expenditures, rather than allowing this investigation to take place as contemplated in the Reorganization of Congress Act.)
3. Against the confirmation of Lilienthal
4. Against Greek-Turkish aid bill
5. For the Taft-Hartley act
6. For cutting foreign relief funds
7. For income tax reduction
8. For increase in rent
9. For price support of wool
10. Against consolidation of housing functions in a single agency.

The correlations presented in Table 1 suggest the presence of more than one common factor. In order to check this, a Spearman single general factor solution was computed, and the residuals examined. The presence of substantial residuals ranging from $- .317$ to $+ .480$ confirmed this suggestion. An attempt was therefore made to extract more than one common factor from this matrix.

The centroid method, as described by Thurstone,⁴ was used to extract three common factors. Since this is a relatively small matrix, it was necessary to repeat the factoring process three times in order to achieve stable communality estimates. For most of these variables, the communality estimates are quite high, and a few exceed 1.00 slightly. This fact raises the question of whether or not this factoring method has failed

⁴ Thurstone, L. L. *Multiple-Factor Analysis*. Chicago: University of Chicago Press, 1947.

to allow adequately for variance attributable to uniqueness in the variable. It may be argued that the opinions expressed on any one issue would be highly reliable in the sense that the expression of the opinion represents a position taken after careful study. In other words, given a similar situation, each Senator would be likely to vote again just as he voted before.⁵ Specificity, the second postulated component of uniqueness, also may be absent from many of these issues, since many of the same determiners may enter into the formulation of opinions on two different issues. Since reliability estimates cannot be made readily for these variables, and since specificity cannot be calculated independently of the estimates of com-

TABLE 2
Obtained Centroids and Communalities

	Centroids			Communality	
	I	II	III	Calculated h ²	Estimates Used in Factoring
P	.906	.337	.225	.99	.99
2	.971	.238	.126	1.00+	.99
3	.943	-.230	-.223	.99	.99
4	.515	-.530	-.050	.55	.57
5	.762	.531	-.387	1.00+	1.00
6	.591	-.641	-.547	1.00+	1.00
7	.914	.328	.022	.94	.95
8	.891	.358	.012	.92	.93
9	.608	-.246	.572	.76	.75
10	.914	-.147	.249	.92	.92

mon-factor and unreliability-factor loadings, it is difficult either to maintain or deny that this method has resulted in a serious over-factoring. The three variables 2, 5, and 6, whose communalities are slightly in excess of 1.00, are of course over-factored. This may not be serious, however, if uniqueness is negligible for these variables.

The three centroids extracted and their accompanying communality estimates are presented in Table 2. The familiar pattern of substantial positive loadings on the first centroid and a bipolar structure for the second and third is apparent here. Since a central reference vector appears here, it is possible to use Thurstone's method of extended vectors.⁶ The plot of the

⁵ The second vote on the income tax reduction Bill (following President Truman's veto) presents some evidence that this is true.

⁶ *Ibid.*, Chapter XI.

extended vectors is given in Figure I. In this figure, the cluster formed by P, 2, 7, and 8—which was obvious in the correlation matrix—is prominent. With these ten variables, all planes are not adequately determined.

Two alternative interpretations are proposed. One would be to regard variables 9, 6, and 2 as each having a maximum

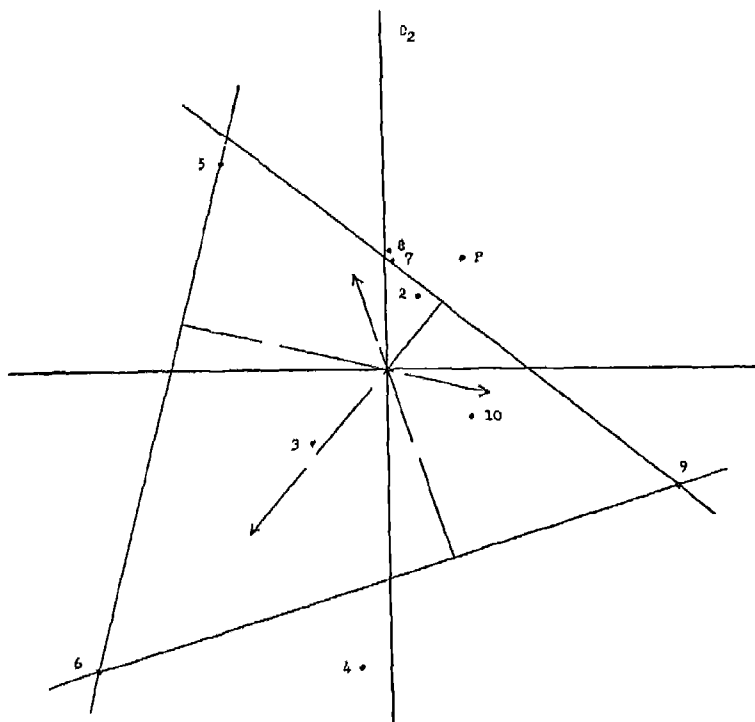


FIG. 1

positive projection on one of the three rotated centroids. One plane would be defined by variables 9, 2, 7, 8, and 5. A second plane would be defined by P, 2, 3, and 6. The third plane is not located precisely. This interpretation would locate one of the rotated centroids in the cluster P, 2, 7, and 8. A schematic interpretation of this factor structure can be given as follows. Plus and minus signs are used to designate the substantial loadings.

		I	II	III	
	P	+			
	2	+			
	7	+			
	8	+			
	3	+	+		
	5	+		-	
	10	+		+	
	9			+	
	4		+	+	
	6		+		

This structure is oblique, with a fairly large correlation between I and III.

A second interpretation consists of regarding variables 9, 6, and 5 as each having a maximum positive projection on one of the three rotated centroids. Variables 9, 2, 7, 8, and 5 would, again, define one plane. The remaining two planes would not be located precisely. A schematic interpretation of this structure is as follows:

		I	II	III	
	5	+			
	8	+		+	
	7	+		+	
	P	+		+	
	2	+		+	
	10	+		+	
	3	+	+	+	
	9			+	
	4		+	+	
	6		+		

This second structure was chosen as the basis of the interpretation to be made. Orthogonal rotation of the obtained centroids gave the factor loadings that appear in Table 3a. These loadings indicate that the structure given above is approximated quite closely by this method of rotation. Table 3b gives the loadings on the three centroids when these reference vectors are permitted to be oblique. As the matrix $\Lambda'\Lambda$ indicates, there are small positive correlations among the three primary reference vectors for this solution. The similarity of loadings for the orthogonal and the oblique solutions is marked, and both solutions suggest essentially the same interpretation.

Rotated Factor I or A has as its highest loadings those for the following three variables.

- 5 Attitude toward the Taft-Hartley act
- 8 Attitude toward increase in rent
- 7 Attitude toward reduction in income tax

TABLE 3a
Orthogonal Rotated Centroids

	I	II	III	IV
5	.999	.000	.063	1.00+
8	.829	-.056	.477	.92
7	.825	-.033	.508	.94
P	.736	-.156	.643	.98
2	.771	-.006	.645	1.00+
10	.467	.209	.808	.92
3	.645	.548	.519	.99
9	.057	.025	.866	.75
4	.101	.574	.453	.54
6	.306	.956	.186	1.00+

TABLE 3b
Oblique Rotated Centroids

	A	B	C
5	.948	.056	-.010
8	.779	-.022	.421
7	.776	.001	.448
P	.615	-.131	.606
2	.700	.021	.583
10	.334	.215	.735
3	.402	.572	.388
9	-.002	.006	.849
4	-.111	.569	.358
6	-.023	.975	.023

MATRIX A'A

	A	B	C
A	1.000		
B	-.260	1.000	
C	-.076	-.174	1.000

It is proposed, tentatively, that this factor represents an attitude toward the interests and methods of big business and management. A substantial loading for P (party designation) on this factor is not surprising. Variable 2 also has a projection of some size on this factor. This issue dealt with governmental organization. This projection may be accounted for in terms of

the well-known opposition to what is often called "bureaucracy" by supporters of the point of view of big business and management. Opposition to Lilienthal (variable 3) also appears to have stemmed in part from disapproval of the TVA, with which he was connected; consequently, the projection of variable 3 on this factor is in keeping with this hypothesis.

Rotated Factor II or B seems to be clearly an "isolationist-internationalist" point-of-view factor. Issues 4 and 6 both dealt with aid to foreign lands. The objections to Lilienthal, it will be recalled from the Congressional hearings, included censure of his foreign ancestry and a questioning of his "Americanism."

Rotated Factor III or C, with its highest loading on variable 9, may be an attitude toward protection of home industries. It is difficult, however, to account for the projection of variable 10 on this factor according to this hypothesis. Another hypothesis that might be advanced is that this factor represents a point of view toward the interests of agriculture as opposed to industry. It should be noted that the alternative interpretation, taking this third factor as bipolar, presents the same problem of naming the factor. Students of political science may wish to bring to bear their greater knowledge of political issues on this problem.

Summary

The analysis of opinions of members of the Senate as expressed in their voting behavior suggests that three points of view underlie these nine specific opinions. Two of these points of view were tentatively identified as: (1) a point of view with respect to the interests and methods of big business and management; (2) a point of view with respect to our relation with foreign countries. The third factor presents an interpretation problem. Further studies should be designed to test certain hypotheses regarding this third point-of-view factor. For example, the addition of variables representing issues that appear to be determined primarily by consideration of protection versus free trade, and of agriculture versus industrial interests, would permit a re-analysis of this correlation matrix that would provide checks on these two hypotheses. Further studies of cor-

relation matrices based on more recent voting records also would be useful in exploring this domain of senatorial attitudes.

The method used in this study represents an application of factorial methods, which were developed primarily in connection with the analysis of mental abilities, to the analysis of opinions. Such an application is, of course, not new. The usefulness of these methods, however, for the student of political science and the student of social psychology may be emphasized by this further application.

•

SOME OBSERVATIONS ON THE COLLEGE BOARD LANGUAGE TESTS¹

HENRY S. DYER

Harvard University

COLLEGES that require the knowledge of a foreign language for the bachelor's degree have the common problem of placing new students in elementary language courses that are appropriate to each student's level of competence. The amount of language study that the student has pursued in secondary school is not ordinarily a very helpful guide. Student A with three years of French in one school may be as much as a year behind Student B who has had only two years of French in another school. Students C and D may both have studied two years of secondary school Spanish, but C resumes it in college after a three months lapse and D after a fifteen months lapse. In any case, when the amount of exposure to language teaching is the sole criterion for course placement, there is considerable likelihood that the entering student of good general competence may find himself in the wrong course.

To avoid as far as possible the misplacement of entering freshmen, Harvard and Radcliffe Colleges regularly require all new students to take placement tests in French, German, and Spanish. When the student enters college, he or she may take up the study of any of these languages at any one of five different levels, or, on the basis of demonstrated competence, he may be exempt from further language study.

The principal, though not the only, objective of the elementary courses is to provide the student with a reading knowledge of the language in question. If he starts from scratch, he ordinarily can acquire this knowledge in four semesters. If he has had previous preparation in the language, he is required to take a test in it either as part of the tests for admission or as

¹The reading tests in French, German and Spanish were supplied by the College Entrance Examination Board.

part of a series of placement tests administered immediately after registration. At the time the data reported in this study were collected, both the entrance and placement tests were those furnished by the College Entrance Examination Board. The results of these tests were the primary basis for placement in the elementary language courses.

It is the purpose of this paper to describe briefly how the results have been used in course placement, and to point out certain features of the tests which have become evident from a study of their validity.

As the numbering of the courses varies from one department to another, arbitrary designations are used in the present report as follows:

A Roman numeral following the name of the course designates its level in terms of semesters. Thus, French I denotes the first semester of beginning French, French III denotes the third semester course in French. A combination of numerals designates an intensive course. Thus, French I II denotes a course that covers the material of French I and French II in a single semester.

I.

In order to make the tests maximally useful for course placement purposes, it is of first importance that the results shall be available *before* the freshman meets his faculty adviser to lay out his course program. At Harvard and Radcliffe the exigencies of freshman registration require that the tests shall be given on a Friday and that the results shall be placed in the hands of students and advisers not later than the following Monday morning. The special techniques by which this task is accomplished have been described elsewhere.² The principal goal of the entire procedure is the preparation of a "Course Placement Recommendation" form, a sample of which is shown in Figure I.

It will be noted from Figure I that the adviser receives not only the scores of his students but specific placement recommendations as well. This practice is followed to relieve the adviser of the necessity of interpreting scores whose significance

² Dyer, Henry S. "Making Test Score Data Effective in the Admission and Course Placement of Harvard Freshmen." *Proceedings of the Research Forum*. Endicott, N. Y.: International Business Machines Corporation, 1946. pp. 56-62.

This flexibility is accomplished by making placement recommendations of two sorts. If the score falls at a point which leaves little doubt about the proper placement, the recommen-

COURSE PLACEMENT RECOMMENDATIONS

Student Jones, John R.		Class 61B	Lesson 275	Date 8/23/47	
Correctly Recommended			Notes	LR Met	Score
Chemistry	B				575
French	V			Yes	610
German	II	III	#		505

FIG. 1. Course Placement Recommendation Form

Both advisers and departmental representatives are supplied with tables showing the Course Placement System for each of the three languages. The data of these tables are summarized

in Table 1. In this table the categories labeled "I or II", "II or III", etc., constitute the borderline zones. These zones represent the areas within which it is estimated from the test score that the student would have received a D or a "weak C" in the lower course if he had taken it at college. It will be noticed that the scores which "place" for any given semester differ from one language to another. These differences arise from the fact that the system for each language depends upon the performance of students who have actually been taking the courses. The necessary information is obtained by administering the tests near the end of each semester to all students who are actually enrolled in the courses. The score systems are, therefore, subject to annual revision.

TABLE 1
Score Systems in Use at Harvard and Radcliffe Colleges for Placement in Elementary Language Courses (September 1947)

Course Recommended	French Scores	German Scores	Spanish Scores
I.	Below 349	Below 429	Below 349
I or II.	350-369	430-439	350-369
II.	370-449	440-499	370-449
II or III.	450-469	500-509	480-499
III.	470-509	510-529	470-509
III or IV.	510-529	530-539	510-529
IV.	530-569	540-549	530-569
IV or beyond.	570-589	550-559	570-589
Beyond IV.	590 and above	560 and above	590 and above

Any comparisons one might be tempted to make from Table 1 as to the relative rates of progress made by students in the different languages is almost certain to be invalid, since the scales on which the scores are reported are not comparable from one language to another. It is possible, for example, that a score of 450 in French represents "more achievement" than a score of 500 in German.

II.

The justification for using any tests for course placement rests on the evidence of their validity. The definition of the validity of a test depends upon the particular way in which it is to be used. If a test is intended to predict how well students will do in college, its validity is defined as the correlation between the test scores and some generally accepted measure of

later performance such as course grades. If a test is intended to provide an index of the student's level of performance at the time he takes the test, its validity is defined as the correlation between the test scores and some generally accepted measure of performance which is applied, or might have been applied, at approximately *the same time that the test itself is applied*. It is the latter definition which is applicable to the present case, for when proficiency in foreign language is ultimately required of *all* the students in college, the problem is not one of predicting the student's future accomplishment; it is a problem of describing his present status.

The importance of this distinction is clear when one considers the case of the new freshman who is trying to decide whether he is ready for second-semester German or third-semester German. If he goes into the second-semester course, he will find it heavily populated with students who have just completed the first-semester course; if he goes into the third-semester course, he will find himself in the company of students who have just moved up from the second-semester course. The problem is to determine whether his knowledge of German is more nearly equivalent to that of the first group of students or to that of the second. The question to be asked of the German test, therefore, is whether and to what extent it can make this determination.

One product of the administrations of the tests in courses has been a considerable accumulation of information concerning the validity of the tests as defined above. Findings from a portion of the data have been reported elsewhere.³ The present paper presents more extensive information. Although the quantity of data available varies somewhat from test to test, there appears to be a sufficient amount to warrant a number of general conclusions.

At the time these tests were given, the results were *not* used in determining the final grades of the students, but it is only fair to state that scores were available to the instructors if they cared to satisfy their curiosity. There is, therefore, a bare

³ Dyer, Henry S. "Validity of the C.E.E.B. Placement Test in French," *The College Board Review*, 1(1947), 1, 12-15, and "Validity of the College Board German Test," *The College Board Review*, 1(1947), 24-26.

possibility that in certain instances the final grade of an individual student may have been affected by the instructor's knowledge of his score. In the opinion of the writer such instances were so rare as to produce no important upward bias in the correlations between the test scores and the final grades.

Table 2 shows the correlations. The probable errors of estimate ($P.E._{est}$) have been given in the table to provide a basis of comparison from course to course. When interpreting the $P.E._{est}$, one should bear in mind that the letter grades were converted to a twelve-point scale as follows: E - = 0, D - = 1, D = 2, D+ = 3, C - = 4, C = 5, C+ = 6, B - = 7, B = 8, B+ = 9, A - = 10, A = 11. Thus, if the $P.E._{est}$ equals 1.0,

TABLE 2
Correlations between GFER Language Test Scores and Final Grades in Various Elementary Language Courses

Course	N	Harvard r	$P.E._{est}$	N	Radcliffe r	$P.E._{est}$
French II...	125	.67	1.7			
French I-II	19	.73	1.3			
French IV	133	.65	1.3	42	.85	.9
Spanish II...	65	.79	1.3	45	.87	1.0
Spanish I-II	15	.65	1.3	6	.66	1.0
Spanish IV	53	.82	1.0	37	.90	.7
German I...	114	.64	1.7	43	.43	2.0
German II	104	.73	1.2			
German I-II*	21	.78	1.3			
German III	52	.78	1.4			
German IV	59	.75	1.3	14	.82	.6
German III-IV	23	.94	.6	7	.85	.9

* Harvard and Radcliffe combined.

the distribution of obtained grades about an estimated grade of C would show a concentration of 50 per cent of such obtained grades within the range C- to C+.

From an inspection of the data in Table 2, it appears that there is a tendency for the relationship between test scores and grades to become larger as the courses become advanced. In other words, the tests seem to be more valid for measuring the achievement of second-year students than they are for measuring the achievement of first-year students. This result is not surprising in view of the fact that the tests were originally designed for use primarily with students who have completed at least two years of language study in secondary school. If the tests

are to be generally used for the course placement of college students, it would be helpful to have them strengthened at the lower end of the scale. It is a question, however, whether such an improvement would be possible in a one-hour test without making sacrifices at other points. A test which provides a sufficiently high ceiling for fourth-semester students must necessarily contain material which is "dead wood" for the first-semester students. Similarly, a fully adequate sampling of the relatively simple material suitable for first-semester students would crowd the test with items incapable of discriminating among more advanced students. Since the placement problem is usually most crucial at the higher levels, the C E.E.B. tests seem to effect a reasonable compromise.

In general, the order of magnitude of the correlations is about as high as one could expect in the circumstances. It may be worth emphasizing that in each case the test score represents a *one-hour* sampling of the student's ability solely to read and understand the language in question. The test provides no opportunity for the student to exhibit his knowledge of "formal grammar," his skill in written translation, the facility with which he can speak or write the language, or his ability to understand the spoken language. Yet the course grades with which the test scores have been correlated reflect, in varying degrees, the student's competence not only in reading, but in these other functions as well. In any case, the tests appear sufficiently valid to warrant their judicious use as aids in course placement.

Two further observations may be made: (1) In five instances out of seven the tests seem to be slightly more valid for the Radcliffe women than for the Harvard men. (2) A rough comparison of the probable errors of estimate found on each test suggests that the three tests are approximately equal in validity

III.

The College Entrance Board tests under discussion are known officially as "The French Reading Test," "The German Reading Test," and "The Spanish Reading Test." The term "reading" in each of the titles raises the question of just what the authors intend. It scarcely needs to be said that reading in

any language is a composite function. A test of "reading" defines the functions in terms of the components it samples, and these components themselves may be defined in terms of the item types used. The item types found in the C.E.E.B. Language Tests are fully described in one of the Board's regular publications.⁴ As defined by the test items, the components are roughly as follows:

- Component A. Recognition of the correct English equivalent of a foreign word.
- Component B. Recognition of the correct foreign equivalent of an English word.
- Component C. Recognition of synonyms among several foreign words.
- Component D. Recognition of correct foreign language usage in context.
- Component E. Recognition of statements in the foreign language which correctly interpret the sense of prose passages in the language.
- Component F. Recognition of English words or phrases as the most exact translations of foreign words or phrases given in context.

None of the tests used in the present study appears to measure all six components, and there is some variability in the extensiveness of the sampling of each. Table 3 summarizes the situation in this respect.

A rough indication of the importance attached to each component in certain of the elementary language courses at Harvard is given by the correlations of the part scores on the tests with the final grades. Table 4 shows these correlations.

Since the individual parts of the test were not separately timed, one cannot be certain that the relatively low correlations for Component E (notably in French II and German II) are not the result of the fact that some students did not allow sufficient time to do themselves justice on this part of the test. Furthermore, the value of any component on one test is not directly comparable to the value of the same component on either of the other two tests because of differences in the number of items used.

It is nevertheless interesting to note that for the tests in all courses Component D (recognition of correct foreign language

⁴ College Entrance Examination Board. *Bulletin of Information*, 1946-1947, 14-17.

usage in context) ranks at or near the top in importance. On the other hand, Component E, (in the case of Spanish, E-F) which comes the closest to duplicating what might be thought of as a "total reading situation" occupies, with one minor exception, the lowest rank in importance. One hesitates to place too much weight on this finding because of the possibility referred to above that variation in the students' rate of work may be rendering these correlations spuriously low. It would

TABLE 3
Number of Items in the CLEB Language Tests Which Sample Various Components

Component	French	German	Spanish
A	50	40	30
B	50	—	30
C	25	40	—
D	30	40	30
E	30	20	30*
F	—	—	30*
Total items	185	140	150

* Combined into a single section of the test

TABLE 4
Correlations of Part Scores with Final Grades

Component	French II (N = 135) r	French IV (N = 144) r	German II (N = 104) r	German IV (N = 89) r	Spanish II (N = 57) r	Spanish IV (N = 52) r
A	.37	.58	.53	.61	66	.63
B	.49	.52			49	.76
C	.46	.37	.58	.65		
D	.49	.52	.64	.66	83	.73
E	.21	.39	.36	.53	*{.57 *.48	
F						

* Components E and F are included in the same part score of the Spanish test.

be worth experimenting with a rearrangement of the tests to investigate this possibility.

Another feature of Table 4 may be remarked upon. In the case of German, the component scores show the same rank order of importance for both second- and fourth-semester courses. In the case of French and Spanish, however, there is some variation in the pattern. For French IV, for instance, Component A (recognition of the correct English equivalent of a foreign word) is at the top with an r of .58, but for French II it is next to the bottom with an r of .37. In Spanish IV, Com-

ponent B (recognition of the correct foreign equivalent of an English word) is at the top with an r of .76 and for Spanish II it is at the bottom with an r of .49. Both of these differences are statistically significant.⁵ They probably reflect some difference in emphasis at the two levels of instruction.

The word "importance" has been used here rather freely. The question "importance for what?" is a legitimate one. We should not be led into the false reasoning that the components have been defined as elements in the reading process and that their relative importance therefore refers to the degree to which they take part in that process. All we may infer is that instructors may weigh certain of the components more heavily than others in their judgment of the worth of a student's performance in the language being taught. The validity of the component (i.e., its importance) is here regarded solely as the degree to which mastery of the skill involved is rewarded by a good grade in the course. A study of the present sort shows how well or ill the test maker has forecast the instructor's judgment. A low correlation may be the result of fault in the test maker's craftsmanship or it may represent a basic difference between what the test maker and the instructor judge to be important. In the latter case, one need hold no brief for either party, but hope that the tangible evidence of disagreement will clarify the thinking of both.

⁵ In each case the difference in the r 's is above the 5 per cent level of confidence.

THE GROUP ORIGINS OF STUDENT LEADERS¹

E. G. WILLIAMSON

University of Minnesota

Critics of the college fraternity system often charge that members are elected or appointed to college-wide leadership offices in numbers that are disproportionate to the relative size of total fraternity membership. In substantiation of such allegations, Mueller and Mueller found that sorority members (largely from homes classified in the higher socio-economic groups) were greatly over-represented in both general extra-curricular activities and in the category of social honors.²

The present study was undertaken in an effort to discover the extent to which fraternity and sorority members dominate the "most important" campus positions in student organizations and activities in the University of Minnesota. Data were gathered for three separate years, 1941, 1944, and 1947, to give comparisons between a pre-war, a mid-war, and a post-war year. The Minnesota study differs from the Muellers' investigation in its restriction to the "most important" activity positions and also with respect to its inclusion of fraternity as well as sorority members in comparison with non-member students.

In such an investigation of the monopolizing of leadership by any particular group or type of students, a most important procedure involves the careful and meaningful classification of positions of leadership in terms of type and importance. The list of important campus activities to be studied was chosen to reflect the general student evaluation of the importance of leadership positions. The classification system used in our study was constructed by common agreement among six staff

¹ Jean Danaher and Donald Hoyt gave significant assistance in the statistical analyses and

² Mueller, Kate Heyner. "Socio-Economic Background and Campus AND PSYCHOLOGICAL MEASUREMENT, III (1943), 143-150.

members of the Student Activities Bureau, a University office designed especially to promote and aid activity groups, five additional professional people working with student groups, two faculty members informed on campus activities, and eight students with extensive knowledge and experience in many different campus groups. With the exception of the sub-category of "queens" (not shown in our tables), only those positions, open equally by regulation to both fraternity and sorority members and to non-members, were included. A further refinement of our classification system was made by selecting a group of "Top Campus Positions," consisting of four sub-groups composed, in general, of the most important positions of the first division. Definitions of the "Important Campus Positions" and the "Top Campus Positions" are given below, together with the number of positions available in 1941, 1944, and 1947, respectively.

IMPORTANT CAMPUS POSITIONS

	Total Number of Leadership Positions		
	1941	1944	1947
1. Governing Boards and Councils: All members of councils responsible to a membership of over 200 students exclusive of any one class or college, such as the Y.W.C.A., etc.	181	154	185
2. Religious and Service Organizations: Presidents and leaders of student church and service groups, such as the Lutheran Students Association, etc.	90	83	107
3. Special Interest Organized Groups: Presidents and leaders of the most active special interest groups, such as the Republican Club, etc.	129	100	154
4. All-University Elected Positions: The entire membership of groups elected by campuswide voting, such as the All-University Student Council	40	37	42
5. Student Publications: The leading members of the editorial and business staffs of the seven publications, such as the <i>Daily</i> (campus newspaper).	34	38	37
6. Activity Honoraries: All members of organizations honoring students for extra-curricular activities, such as, Mortar Board.	171	154	200

TOP CAMPUS POSITIONS

	Total Number of Leadership Positions		
	1941	1944	1947
1. Chairmen:—All appointments to major chairmanships for important campus-wide positions, such as, Homecoming Chairman.	42	45	58
2. Major Campus Posts:—Presidents of groups 1, 4, and 6; editors of Group 5; and major chairmen.	60	59	68
3. Most Important Positions:—Presidents of Groups 1 and 4 and editors of the three main publications.	14	15	15
4. Heads of all Organizations:—Presidents of university recognized activities or organizations.	174	140	180

Fraternal groups are classified by the University as academic or professional. The first includes all fraternities and sororities who hold membership open to students, regardless of the field or college in which they are studying. The second type limits membership to students studying in certain specified fields. Aside from this, the groups are essentially alike, both in purpose and in function. But the two groups were treated separately in this study for purposes of comparison. Fraternities and sororities were classified jointly as either academic or professional fraternal groups. Combining them thusly is statistically justified when percentages of enrollment and of participation are compared for the three years, no difference being significant at the .01 level of probability.

Table 1 below shows the distribution of fraternal membership, by per cent, in each of the important campus activities and in the total student enrollment for the three years considered.³ The most obvious generalization is that Table 1 (and Table 2) reveal a marked over-representation on the part of fraternal groups in these activities. A separate study of sorority members, not shown in this paper, reveals a degree of concentration similar to that found by the Muellers.

When the total number of available leadership positions is divided into the total number held by academic fraternal, pro-

³ Total enrollment for the three years in question was 13,603, 7,130; and 26,293, respectively. All data were taken from official University records for the Winter Quarters of 1941, 1944, and 1947.

fessional fraternal, and non-fraternal members, the total per cent of representation is obtained. (See figures at the bottom of each column in Table 1.) When these percentages are compared with the percentage of the total student enrollment for each of the groups, (see the first, top row of Table 1), critical ratios significant at a level of probability far beyond .01 were

TABLE 1
Distribution of Membership, by Per Cent, in Important Campus Activities and in the Total Enrollment for 1941, 1944 and 1947

Group Classification	1941			1944			1947		
	Fraternal	Acad	Prof	Fraternal	Acad	Prof	Fraternal	Acad	Prof
Total Enrollment	17.0	19.4	22.6	18.4	11.6	68.6	9.1	5.3	85.6
Gov. Bds. & Councils	40.9	27.1	32.0	44.8	30.5	24.7	40.5	38.8	28.7
Relig. & Service	23.3	26.7	50.0	24.1	23.5	55.4	21.5	16.8	61.7
Special Interest	22.5	23.2	54.3	31.1	19.0	50.0	25.3	14.3	60.4
All-U. Elected	40.5	30.0	22.5	60.0	10.8	21.0	61.0	26.2	11.9
Publications	29.4	41.2	29.4	44.7	34.2	21.1	21.0	29.7	48.7
Activity Honoraries	43.9	22.2	33.9	40.9	23.4	35.7	38.5	24.0	37.5
Totals	35.1	28.7	36.0	39.9	26.7	33.4	33.8	25.0	40.3

TABLE 2
Indices of Fraternal Representation in Leadership Positions

Group Classification	1941			1944			1947		
	Acad Fr	Prof Fr	Non Fr	Acad Fr	Prof Fr	Non Fr	Acad Fr	Prof Fr	Non Fr
Gov. Bds. & Councils	241	261	44	243	235	36	445	581	34
Relig. & Service	137	257	69	131	158	81	216	317	72
Special Interest	142	223	75	168	146	73	278	270	71
All-U. Elected	279	288	31	167	83	31	680	494	14
Publications	173	396	49	243	263	31	237	560	57
Activity Honoraries	258	213	47	222	189	52	423	453	44
Totals	268	276	50	212	265	49	371	489	47
Chairmen	368	91	52	217	171	55	379	162	66
Major Campus Posts	176	416	37	184	208	57	388	583	39
Most Important Posts	84	412	59	217	266	39	366	755	31
Heads of all organizations	271	271	36	241	192	45	464	504	36
Totals	248	282	39	222	196	49	429	470	42

obtained in every case. Thus the possibility that these are variations due to chance is statistically eliminated. We see in Table 1 that fraternity students constituted 17.0 per cent of the total student enrollment in 1941 but these same students held a total of 35.3 per cent of the most important leadership positions. Professional fraternity members constituted 10.4 per

cent of the enrollment and held 28.7 per cent of the positions. Conversely, non-fraternal students made up 72.6 per cent of the enrollment but held only 36 per cent of leadership positions.

One of the most striking features of Table 1 is the showing made by fraternal groups in 1947. As the table indicates, the academic organizations composed but 9.1 per cent of the total student enrollment and the professional groups but 5.3 per cent. Yet these two groups held nearly as high a percentage of important positions in 1947 as in the other two years when their membership made up a much larger proportion of the student body. In fact, when the total representation by per cent is placed on an index number basis, as in Table 2, with 100 representing proportionate representation according to the composition of the entire student population, the following figures are obtained for the three years in chronological order: Academic fraternal groups—208, 217, 371; Professional fraternal groups—276, 205, 489; Non-fraternal groups—50, 49, 47. That is, fraternal students hold a number of leadership positions greater than their combined proportionate size of the student body. This fact may hold important implications for the new student, especially during such periods as were experienced after the war, when the student enrollment expanded rapidly while fraternal groups admitted only a few new members. It would seem that the new student would have not only a small chance of attaining fraternal affiliation, but also an even smaller chance of gaining an important campus position.

An analysis of the representation (Tables 1 and 2) in each of the groups of important positions yields other points of interest. It is not the purpose of this paper to offer explanations for the figures presented or the trends cited, but rather to note their existence and to suggest a few of the possible implications. Representation in the religious and service and special interest groups (see Table 1) is quite evenly distributed between fraternal and non-fraternal members, though a degree of overrepresentation on the part of both academic and professional fraternal organizations is apparent for each of the years. When the index scheme is applied here (see Table 2), the religious and service groups show academic fraternities and sororities with index numbers of 137, 131, and 236 for 1941, 1944, and

1947 respectively. Professional groups have representation indices of 257, 158, and 317; while the non-fraternal members are represented with indices of 69, 81, and 72. Very similar figures result when special interest representation is placed on the index system. These are the highest representation figures attained by the non-fraternal groups, emphasizing the universal concentration of fraternities and sororities.

The largest representation of academic fraternal organizations occurs in the Governing Boards and Councils, All University Elected, and Activity Honoraries. Professional groups find their greatest concentration in Publications, though they are also markedly dominant in the three categories given for the academic organizations. Their added prestige in the Publication group is explained by the classification of Sigma Delta Chi, the professional journalism fraternity, members of which held over 20 per cent of all important publication positions for each of the three years.

Index numbers on participation of academic organizations increase in size each year in the classifications of Governing Boards and Councils, Special Interest, All University Elected, and in total representation. The indices of non-fraternal groups for the same categories show a decrease each year, while the professional organizations, in every case, obtained their highest representation in 1947 and their lowest in 1944.

Insofar as the figures obtained for the first, middle, and last year can be considered representative of a span of seven years, it would seem that the degree or extent of concentration of fraternal groups in leadership positions is increasing, and that of non-fraternal groups is decreasing. As was pointed out earlier, however, this may be a function of the size of the total enrollment. What is needed is a study of many more years with varying University populations to verify or disprove the trends indicated.

Table 3 is a presentation of the same type of data as was given in Table 1, except that the more select groups—the Top Campus Positions are considered. An examination of the percentages given in the Total column indicates even heavier concentration of fraternity and sorority members in these positions.

By again resorting to the index scheme, (see bottom section of Table 2) several interesting facts are uncovered. In 1941, academic fraternal groups obtained an index number of 84 in the Most Important Posts classification. Also in 1941, professional organizations were under-represented in the Chairmen category, as indicated by an index number of 91. These two instances are the only ones in this study in which fraternal groups were shown to be under participating, proportionate to their size of membership. In both of these particular cases, there was a large over-representation of the other fraternal groups, so that the non-fraternal members attained index numbers of only 59 and 52 for Most Important Posts and Chairmen respectively for that year, 1941.

TABLE 3

Distribution of Membership, by Per Cent, in Top Campus Positions and in the Total Enrollment for 1941, 1944, and 1947

Group Classification	Acad. Fr.	1941 Prof. Fr.	Non- Fr.	Acad. Fr.	1944 Prof. Fr.	Non Fr.	Acad. Fr.	1947 Prof. Fr.	Non- Fr.
Total Enrollment	17.0	10.4	72.6	18.4	13.0	68.6	9.1	5.3	85.6
Chairmen . . .	52.4	9.5	38.1	40.0	22.2	37.8	34.5	8.6	56.9
Major Campus Posts	39.0	43.3	26.7	33.9	27.1	39.0	35.3	30.9	33.8
Most Important Posts	14.3	42.8	42.9	40.0	33.3	26.7	33.3	40.0	26.7
Heads of All Orgns.	46.0	28.2	25.8	44.3	25.0	20.7	42.2	26.7	31.1
Totals	42.1	29.3	28.6	40.9	25.5	33.6	39.0	24.9	36.1

Certain trends are again made apparent by means of the index method. Over the period studied, academic fraternal members gained increasingly larger indices in Major Campus Posts and in Most Important Posts, while non-fraternal members showed a steady increase in participation in chairmanships and a decrease in Most Important Posts. Professional groups followed the same pattern as that indicated in Table 1, except that they were most frequently represented in the Chairmen category in 1944 and least frequently represented in that category in 1941.

It is of particular interest to note that again both types of fraternal groups were most highly represented in 1947 and had their least representation in 1944—the year that non-fraternal members made their best showing. Again the size of enroll-

ment may be involved as a causal factor in ways not readily apparent

The effect of the war is well illustrated when a comparison is made between the percentages of fraternal members holding important campus positions in the three separate years. For this consideration, the critical ratios between the percentage of membership participating in the important activities in 1941 were compared with the corresponding percentage in each of the other years; the figures of 1944 were also compared with those of 1947.

Critical ratios of over 5.0 (not shown in this paper) were obtained in the consideration of academic fraternal groups when 1944 was compared separately with 1941 and 1947. These ratios were approximately 2.0 when professional groups were considered, and were again over 5.0 when the two fraternal groups were combined. However, when percentage differences between 1941 and 1947 were considered, and critical ratios computed, the highest ratio obtained was .51. Thus, the similarity between fraternal participation by per cent of membership for 1941 and 1947 is statistically established, while true differences are found in every instance of comparison with 1944 of each of the other two years.

Without additional data for other years, it is not possible to establish 1941 and 1947 as representative years with respect to activity participation. Nor is it possible to establish that the year 1944 is significant for its under-concentration of fraternal members in leadership positions. Whatever the causal factors involved, fraternal participation in important activities was strikingly similar during a pre-war and a post-war year, and was strikingly dissimilar between these years and a mid-war year.

But the preceding discussion of concentration of leadership among fraternal groups does not tell the entire story. On the assumption that a selected few fraternal groups may dominate leadership activities to a considerable extent, in other words that there will be concentration within concentration, an analysis was made of the number of important positions each individual fraternal organization held. The results indicate that in 1941 ten academic fraternities and sororities having 27.6

per cent of the total academic fraternal enrollment, held 58.9 per cent of the important positions filled by academic fraternal students. In 1944, nine of these ten groups (one group having become inactive) having 34.4 per cent of the total academic fraternal enrollment, held 62.2 per cent of the important positions filled by academic fraternal members. In 1947 these same nine groups representing only 19.4 per cent of the academic fraternal membership, held 48.9 per cent of the positions filled by academic fraternal members.

The corresponding figures for the professional groups are even more significant. Four individual groups were considered in each of the three years, with the following percentages (combined for the four groups) of professional fraternal enrollment and leadership positions held. In 1941 four professional groups with 18.7 per cent of the total professional fraternal membership, held 72.1 per cent of the positions of leadership. By 1944 members of these same four groups with 13.3 per cent of the membership occupied only 62.7 per cent of the leadership positions. In 1947 the corresponding percentages had decreased to 11.9 and 50.7, respectively. Future studies will determine whether this downward trend will continue.

The fact of concentration of fraternal students in positions of activity leadership seems to be well established. But the methods used to achieve such concentration, as well as the educational and social significance and implications are not readily apparent. More exhaustive investigations are needed to determine the significance of several possible factors. In the first place, speaking from the viewpoint of an objective but not unfriendly researcher, these data might prove to be evidence of the validity of the contention of fraternal leaders that: (1) fraternities and sororities select freshmen who are "potential leaders"; and (2) fraternities "train" leaders. But the data could also equally prove to be support for those unfriendly critics who contend that: (1) fraternities use student political parties and techniques to "get" their members elected and appointed to positions of leadership; and (2) fraternities make members of those non-member students who have achieved positions of leadership *after* they have so achieved.

The determination of which of these four, or other, possible

interpretations of our facts of concentration of fraternal members in important positions of leadership, is a matter of considerable importance to one who has an interest in using scientific methods to identify and study the processes used by student group members to select leaders. These interpretations also have serious importance for one who searches for valid ways in which to identify potential leaders and to insure their election, by student groups, to leadership responsibilities. If fraternities and sororities have perfected socially desirable and effective methods of identifying, training, and electing effective leaders, then group workers by all means should learn from such accumulated experiences. It is for these reasons that we undertook this study of the origin of activity leadership. The possible application of techniques perfected in fraternal groups to non-fraternal members does not in any way imply a necessary criticism of fraternities. Moreover, such a broader application will not necessarily weaken the fraternity system, unless the techniques so identified and used would prove to be anti social in their fundamental character. But if the techniques are sound and educational and socially desirable, then their use will add to the strength of the fraternal system and, indeed, may well add to the membership and number of fraternal groups.

The author's interest is based upon a search for effective means of aiding more and more students, be they fraternal or non-fraternal, to gain through college activities those enriching personal experiences which are afforded by student activities, within the potentialities of the individual student. More effective utilization of activities, by students *and* personnel workers is the purpose of this first preliminary exploration of the origins, techniques and instrumentalities involved in the development of effective student leadership.

THE DEPENDENCE OF FACTORIAL COMPOSITION OF APTITUDE TESTS UPON POPULATION DIFFERENCES AMONG PILOT TRAINEES I. THE ISOLATION OF FACTORS¹

FRANK J. DUDEK
Northwestern University

Introduction

THE purpose of this study was to determine the comparability of factors resulting from a centroid analysis of a battery of tests administered to groups of pilot trainees representing different selective conditions. Various theories have been proposed to account for the relationships among psychological tests. One of the most promising has been that of weighted group factors in linear combination. In essence all methods of factor analysis attempt to determine the number of factors which must be postulated to account for the relationships observed among tests, and to determine the loading or weight of each test in each factor. Thurstone's centroid method (18), for example, provides a technique for extracting independent reference axes. If a group-factor theory is accepted as a convenient working hypothesis it becomes important next to examine the common factors isolated and to determine their natures and relationships.

Numerous studies have been made of the common factors to be found in various batteries of tests (2, 3, 5, 9, 15, 20, 23, 24, 25, 26). At the same time there has been considerable interest in the various conditions that influence factor patterns. Guilford (11) found that factors resulting from an analysis of the ten subtests of Seashore's pitch discrimination test were related to the difficulty of the items. Several investigators

¹This article is based on a dissertation submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy at the University of Southern California. The writer wishes to express his sincere appreciation to Dr. J. P. Guilford for valuable assistance while the study was in progress.

have shown that age may have some influence on factor patterns (1, 4, 8, 14, 15, 22)

Most pertinent to the present study are the effects of selection of samples upon factor patterns. This problem, however, has not been studied from an empirical standpoint. Analytical studies (16, 17) have indicated what might be expected when selection occurs. Thurstone (21) studied the problem by using measures of boxes as variables and imposing various conditions of selection on his samples. He showed that the physical interpretation of primary factors remained the same under wide variations in selection of both subjects and tests, and that simple structure was invariant under widely different selective conditions. The present study may be considered, then, as an attempt to test empirically the hypothesis that similar primary factors will be obtained from a battery of tests administered to different samples representing different selective conditions.

Tests

The test battery analyzed was the *AAF November 1943 Classification Battery*. The tests are briefly described below (for complete descriptions see Guilford, 12). The tests are numbered as they appear in subsequent intercorrelation and factor matrices.

1. *Biographical Data Blank Navigator* (CE602D). This blank requested data regarding the personal history of the subject. It consisted of 65 items about personal history, academic achievement, activities and hobbies, and work experience. Items keyed were determined by item-analyses which utilized success in navigator training as a criterion.

2. *Biographical Data Blank Pilot* (CE602D). This blank was the same as the first blank described. Items keyed to obtain this score, however, were based on item-analysis data which utilized success in pilot training as a criterion.

3. *Spatial Orientation I* (CP501B). This test was composed of 48 items that required the individual to match small aerial photographs with the corresponding portion of a large photograph.

4. *Spatial Orientation II* (CP503B). This test was comprised of 48 items that required the subject to match an aerial photograph with the area on a map which corresponded to the photograph.

5. *Reading Comprehension* (CI614H). This test contained 36 items regarding seven paragraphs of reading material. The items about each paragraph followed it directly in the booklet and these were answered before the next paragraph was read.

6. *Dial and Table Reading Test* (CP622-621A). In the first part of this test the subject was required to read indications on dials which were calibrated in different units. This part consisted of 60 items. The second part, consisting of 86 items, required that the individual find values within a large table of figures.

7. *Mechanical Principles* (CI903B). This was a test of 40 items pertaining to mechanical principles and devices.

8. *Instrument Comprehension I* (CI615B). The task in this test was to interpret five dial readings on an instrument panel and reconstruct the attitude of a plane they supposedly represented, and then choose, from a series of five possible responses, the verbal description that properly described the plane. The test consisted of 15 items.

9. *Instrument Comprehension II* (CI616B). In this test two indicators represented the flight attitude of a plane. The task was to find the picture of a plane which was in the proper flight attitude as indicated by the two instruments. Sixty items comprised the test.

10. *General Information Test* (CE505E). This test consisted of 100 items measuring general information about planes and flying. Also included were items requiring mechanical information and some pertaining to athletics.

11. *Mathematics A* (CI7021). This test was composed of 35 items primarily in the form of equations employing algebraic symbols.

12. *Mathematics B* (CI206C). This test consisted of 30 items of the "story problem" type.

The tests described above were paper-and-pencil tests and might be designated "group" tests. They were administered to large groups of individuals by a trained examiner assisted by trained proctors. Items were multiple-choice and answers were made on answer sheets which were machine scored. All tests were timed. Tests described below, which comprised the remainder of the battery, were designated "psychomotor tests." These were administered to small groups of subjects (about 6 per group) by a trained examiner. Each test required performance on apparatus. The paper-and-pencil tests were given during one day of testing and the psychomotor tests were given on the next testing day.

13. *Rotary Pursuit Test* (CP410B). This test required the subject to follow a revolving target with a stylus. The ap-

paratus resembled a phonograph turn-table near the outer edge of which was embedded a small brass disc. The turn-table revolved at a constant rate of 60 RPM and the task was to keep a metal stylus in contact with the disc as it travelled around. Contact between the stylus and disc activated a clock and the score was the length of time in a given period that contact was made. The task was made complex by requiring the subject to keep one of two buttons depressed with his left hand. Time of stylus-on-disc was counted only when the correct button was depressed.

14. *Two-hand Coordination Test (CM101A)*. This test required that both hands be used simultaneously in operating two lathe-type handles to follow a target moving in an irregular path. The target was a disc in a revolving turn-table, but various grooves and cams caused the disc to move in an irregular path rather than in a circle as in the *Rotary Pursuit Test*. The stylus, which was to be kept in contact with the target, could be moved by appropriate manipulation of the two lathe-type handles. One handle moved it forward or backward, and the other moved it to the right or left. The handles, also, were operated in planes that were at right angles to one another. Scoring was in terms of time that the stylus was in contact with the target.

15. *Complex Coordination Test (CM701A)*. In this test the subject sat facing a panel which contained paired rows and columns of lights. A pattern of lights appeared on the panel and the task was to reproduce this pattern by appropriate movements of a "stick" and rudder-bar. When the pattern of lights was successfully reproduced, it would change automatically and a new movement was called for on the part of the subject. Scoring was in terms of the number of correct reproductions accomplished in a given period of time.

16. *Rudder Control Test (CM120B)*. In this test the subject sat in a mock cockpit that was rather unstable. He sighted over the cockpit at a target a small distance away. The task was to control foot pressures on rudder-bars so as to keep the cockpit in line with the target. Any change in pressure of the two feet, once the cockpit was in line with the target, would cause the cockpit to turn away from the target in the direction of greater pressure. Scoring was done with automatic timers which registered the amount of time in a given period that the cockpit was in proper alignment with the target.

17. *Discrimination Reaction Time Test (CP611D)*. In this test the subject sat facing a vertical panel on which it was possible to flash red and green lights in different relations to one another. In front of the subject was a horizontal panel of switches. Which switch the subject was to push depended on the relative position of the red to the green light. Scoring was in terms of total time to make correct discriminations and responses in a given number of trials.

18. *Finger Dexterity Test* (CM116A) In this test the subject sat at a table on which was a board containing 48 pegs. The task was to lift, turn, and replace the pegs in the board. Scoring was in terms of the number of pegs lifted, turned, and replaced in a given interval of time. (For detailed descriptions of the psychomotor tests see Melton (13)).

19. *Pilot Criterion.* The pilot criterion is the variable against which all of the classification tests described above were validated. It is an artificially dichotomous criterion which consisted of information as to whether the trainee passed or failed at any of the progressive stages of pilot training. Correlations between it and any of the remaining variables are biserial coefficients of correlation.

The Samples Studied

The Restricted Group.—After an individual made application for aircrew training he took a qualifying examination. Only those individuals who made a prescribed mark on this test were sent on to classification centers where they took the battery of classification tests. This qualifying, or "screening," examination eliminated from 40 to 60 per cent of the applicants. Thus, the sample of individuals who took the classification tests was curtailed. Factorial studies on samples tested at classification centers were done on this kind of restricted sample. This constitutes one of the groups under consideration and has been designated the "Restricted group." Data concerning factor loadings for this group were simply taken from their original source (12), as will be described later.

The Experimental Group.—This group consisted of men who were admitted to training without regard for scores on the qualifying examination. This was done as a project designed to study selection and classification procedures. From the standpoint of the present study it is a significant sample because it was not restricted on a basis of either the qualifying examination or the classification battery. This group is described in detail by Flanagan (7).

The Wasp Group.—The WASP (Women's Auxiliary Service Pilot) group consisted of women who were in training to learn to fly army-type aircraft. No screening examination like that given to prospective aircrew members was used for their selection, but students had to present at least 35 hours of previous flying as a prerequisite for admission to training. The average

age of this group was 24 years. Twenty five per cent of the group had completed four or more years of college and 40 per cent had had some college training. Only 3 per cent had not completed high school.

Comparison of Groups—One basis for comparison is the proportion of each group above the qualifying score of the screening examination. Of the group restricted on a basis of this score, of course, all have scores at or above this qualifying mark. Of the Experimental (unrestricted) group 42 per cent fell below

TABLE 1
*Means and Standard Deviations for the Experimental Samples on Tests Comprising the AIF
N-Order to 42 Classification Battery*

Test	Experimental Group (N = 177)		Restricted Group (N = 191)		WASP Group (N = 195)	
	M	SD	M	SD	M	SD
1. Biographical Data—Navigator	22.19	3.31	21.79	3.73	21.73	3.14
2. Biographical Data—Pilot	20.29	6.96	20.76	6.47	21.45	5.74
3. Spatial Orientation I	27.69	5.75	26.33	5.61	29.61	6.91
4. Spatial Orientation II	16.96	5.12	19.72	6.44	21.25	6.95
5. Reading Comprehension	11.83	12.53	14.17	11.86	21.43	14.21
6. Dial and Table Reading	27.53	11.13	32.56	9.85	35.64	9.63
7. Mechanical Principle	20.15	9.10	8.49	8.96	21.02	6.52
8. Instrument Comprehension I	7.32	3.40	8.30	3.38	10.76	3.21
9. Instrument Comprehension II	24.69	13.92	26.69	16.71	26.44	11.66
10. General Information	31.48	15.28	32.33	13.40	42.19	11.26
11. Mathematics—A	4.50	6.80	5.29	6.73	5.97	6.95
12. Mathematics—B	6.93	9.12	10.31	8.91	11.47	9.25
13. Rotary Per cent	48.97	10.28	49.19	9.97	44.37*	9.37
14. Two hand Coordination	49.28	10.22	50.53	10.16	38.93*	10.20
15. Complex Coordination	45.99	10.91	49.55	9.49	46.33*	11.56
16. Rudder Control	43.73	12.96	49.45	10.11	61.86*	5.73
17. Discrimination Reaction Time	45.61	12.22	49.82	10.08	44.44*	12.65
18. Finger Dexterity	60.65	10.52	51.26	10.44	47.85*	9.64

* N = 191.

this qualifying score. Data are not available for the specific sample of WASP's studied in this problem, but studies on larger groups (28) indicate that about 30 per cent of the WASP group had scores lower than the minimum qualifying score for the men.

A more pertinent basis for comparison involves the means and SD's of the groups in the tests being analyzed. These are shown in Table 1. Inspection of Table 1 reveals several significant comparisons:

1) The Restricted group has higher means than does the

Experimental group for all tests except the Biographical Data Tests.

2) The SD's for the Restricted group are smaller than for the Experimental group for all tests except *Biographical Data—Navigator*, and here the difference is negligible.

3) The WASP's tend to be intermediate to the men's samples with respect to means and SD's. They excel on some tests and are inferior in others. Tests in which they excel seem to be related to educational background, (Reading Comprehension, General Information, Mathematics); they tend to do more poorly than the men on the apparatus tests. An exception is the *Rudder Control Test*, which may be the result of their experience in piloting planes and manipulating rudder pedals.

There seems little doubt that actual curtailment resulted from selecting trainees on a basis of the qualifying examination and classification battery. The extent is not definitely known but it seems logical to assume that curtailment must also be in whatever factors are measured by the various tests of the battery. When the Restricted and Experimental samples are compared for Pilot Stanine, which is an index from one to nine based on a weighted average of all tests in the battery, significant differences are obtained with respect to means and SD's.² It seems relatively certain, then, that the three samples do represent different conditions of selection. Whether these different conditions of selection will alter the factors isolated from the battery of tests is one of the primary considerations of this study.

Intercorrelations of Tests

Table 2 presents the r 's among the test variables for the Experimental group. These r 's appear in DuBois (6, Table 5.9) as part of a larger matrix. The last row (variable 19) contains the biserial r 's between the pass-fail pilot criterion and each of the 18 test variables. All other r 's are product-moment coefficients. The r 's are based on 1012 cases, while the means and SD's reported in Table 1 are based on 1275 cases. The discrepancy is due to the fact that the r 's are only for that sam-

² Experimental Group: Mean = 4.08; SD = 2.06
Restricted Group: Mean = 4.68; SD = 1.76

TABLE 2
Intercorrelations of Test Variables and Pict Criterion for the Experimental Group

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. BDN	18*																		
2. BDP	11	10																	
3. SOI	23	23	1																
4. SOII	23	23	43	31															
5. R. Comp	16	11	43	31	43														
6. DR-TR	-55	39	15	31	43	43													
7. MP	52	14	35	31	43	43	43												
8. ICI	55	25	35	43	43	43	43	43											
9. ICH	54	45	35	43	43	43	43	43	43										
10. GI	13	58	24	35	43	43	43	43	43	43									
11. Math A	13	53	16	35	43	43	43	43	43	43	43								
12. Math B	51	17	12	35	43	43	43	43	43	43	43	43							
13. RP	51	28	35	35	43	43	43	43	43	43	43	43	43						
14. ARC	51	23	35	35	43	43	43	43	43	43	43	43	43	43					
15. CC	51	24	35	35	43	43	43	43	43	43	43	43	43	43	43				
16. RC	15	12	31	35	43	43	43	43	43	43	43	43	43	43	43	43			
17. DRT	57	25	35	35	43	43	43	43	43	43	43	43	43	43	43	43	43		
18. FD	11	34	31	35	43	43	43	43	43	43	43	43	43	43	43	43	43	43	
19. PC																			

* Decimal points have been omitted.

ple that had records available for flight training. From the time of testing until the time of graduation (or elimination) from training some records became unavailable. However, there is no reason to believe that the records lost to the study biased the sample in any way.

All but two of the r 's in Table 2 are positive. The two negative coefficients do not differ from zero appreciably (standard error of an $r = .00$ for a sample of this size is .03).

Table 3 presents the r 's among the variables for the WASP sample.³ This table has seven negative entries. Of these, five involve the *Finger Dexterity Test*. None are significantly different from zero. (Standard error of an $r = .00$ for a sample this size is .07)

Intercorrelations are not presented for the Restricted group since these data were not analyzed as a part of this study. Tables of r 's for this group and factor analysis results appear in Guilford (12). Pertinent data have been included in this study for purposes of comparing factor loadings of tests.

Factorial Analyses

The matrices of intercorrelations presented in Tables 2 and 3 were factor analyzed by Thurstone's centroid method. The procedures followed for a solution were as described by Guilford (10).

One of the criticisms of the centroid method has been that there are no criteria as to how many factors to extract from a given matrix of r 's. The indeterminacy of the number of factors to extract is due in part to the fact that the standard errors of the factor loadings cannot be computed. The number of factors, then, is left to logical considerations rather than to a mathematical rationale.

One argument is that more factors should be extracted than will ultimately be used. Since the important consideration is to find the meaningful common factors, if extraction of factors is stopped too soon important amounts of the communalities among the tests may be left unaccounted for. When more factors are extracted than are needed, however, those factors

³ The means, SD's, and r 's for the WASP group were furnished by Col. Walter F. Deemer.

TABLE A

Intermedian μ of Test Variables and Percentiles of the Normal Curve $V = 100$

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. BDN....	—	215	221	227	231	235	239	243	247	251	255	259	263	267	271	275	279	283	287
2. BDP....	215	—	221	227	231	235	239	243	247	251	255	259	263	267	271	275	279	283	287
3. SO I	215	221	—	227	231	235	239	243	247	251	255	259	263	267	271	275	279	283	287
4. SO II	215	221	227	—	231	235	239	243	247	251	255	259	263	267	271	275	279	283	287
5. R. Comp	215	221	227	231	—	235	239	243	247	251	255	259	263	267	271	275	279	283	287
6. DR-TR	215	221	227	231	235	—	239	243	247	251	255	259	263	267	271	275	279	283	287
7. MP	215	221	227	231	235	239	—	243	247	251	255	259	263	267	271	275	279	283	287
8. IC I	215	221	227	231	235	239	243	—	247	251	255	259	263	267	271	275	279	283	287
9. IC II	215	221	227	231	235	239	243	247	—	251	255	259	263	267	271	275	279	283	287
10. GI	215	221	227	231	235	239	243	247	251	—	255	259	263	267	271	275	279	283	287
11. Math A	215	221	227	231	235	239	243	247	251	255	—	259	263	267	271	275	279	283	287
12. Math B	215	221	227	231	235	239	243	247	251	255	259	—	263	267	271	275	279	283	287
13. RP	215	221	227	231	235	239	243	247	251	255	259	263	—	267	271	275	279	283	287
14. -HC	215	221	227	231	235	239	243	247	251	255	259	263	267	—	271	275	279	283	287
15. CC	215	221	227	231	235	239	243	247	251	255	259	263	267	271	—	275	279	283	287
16. RC	215	221	227	231	235	239	243	247	251	255	259	263	267	271	275	—	279	283	287
17. DRT	215	221	227	231	235	239	243	247	251	255	259	263	267	271	275	279	—	283	287
18. FD	215	221	227	231	235	239	243	247	251	255	259	263	267	271	275	279	283	—	287
19. PC...	215	221	227	231	235	239	243	247	251	255	259	263	267	271	275	279	283	287	—

* Decimal points have been omitted.

which are in truth residuals cannot ordinarily be rotated into meaningful positions and the amount of variance attributable to them will be small--and they can be dropped from consideration. According to this logic the danger is in extracting too few factors rather than too many. Since the superfluous factors will show up as residuals or meaningless artifacts they will not influence the interpretation of the primary factors, but too few factors would be difficult to rotate to meaningful positions and would leave important amounts of communalities unaccounted for.

The criterion used in this study to determine the number of factors to extract was the contribution of the factors to the intercorrelations of tests. The r between two variables may be expressed as the algebraic sum of the cross-products of loadings in factors common to the two variables. Thus, the maximum contribution which a factor makes to any r can be determined simply by computing the product of the two highest factor loadings in this factor. It is apparent that when this product is less than the standard error of the corresponding r , the maximum contribution which this factor is making to the r is certainly no greater than chance. Any factor which does not meet this criterion cannot be considered a common factor because it does not contribute to the relationship between any two variables an amount that could be considered significant.

The centroid factors and the estimated and computed communalities of tests determined from the r 's of the variables for the Experimental group are presented in Table 4. Similar data for the WASP group is contained in Table 5.

Eight centroid factors were extracted from the Experimental group matrix. Examination of the loadings for the eighth factor reveals that the largest loadings are .159 and .176 for variables 1 and 2 respectively. The product of these loadings is .0279. The standard error of the original r between these tests is .03. According to our criterion, then, no more factors were extracted. Examination of the residuals showed but one higher than .04 and this also would indicate that not much more variance would be involved.

Seven centroid factors were extracted from the WASP group

TABLE 4
Centroid Factor Loadings, Estimated and Computed Communalities of Variables for
Experimental Group

Variable	I	II	III	IV	V	VI	VII	VIII	h _e	h _c
1. BDN	.152*	.112	.076	.076	.36	.17	.44	.159	.195	.228
2. BDP	.398	-.286	.241	.270	.279	.171	.110	.176	.375	.461
3. SO I	.529	.160	.271	.249	-.06	.279	.079	.128	.412	.445
4. SO II	.595	.145	.046	.271	.121	.31	.145	.121	.614	.511
5. R Comp	.574	.277	.227	.172	.174	.142	.174	.170	.679	.521
6. DR-TR	.658	.379	.179	.079	.072	.077	.079	-.022	.632	.616
7. MP	.617	-.166	.212	.195	.247	.249	.145	.14	.699	.610
8. IC I	.626	.249	.194	.172	.198	.178	.079	.07	.677	.546
9. IC II	.663	.046	.073	.19	.279	.077	.077	.051	.610	.517
10. GI	.614	.172	.329	.077	.071	.211	.074	.044	.598	.592
11. Math A	.467	.379	.275	.257	.277	.072	.077	-.022	.482	.482
12. Math B	.487	.395	.162	.127	.077	.178	.075	.042	.561	.550
13. RP	.387	-.327	.213	.077	.079	.071	.122	.079	.328	.341
14. 2HC	.544	-.246	.191	.172	.071	.077	.212	.051	.555	.547
15. CC	.624	-.224	.292	.072	.072	.11	.079	.171	.518	.551
16. RC	.395	-.448	.079	.076	.078	.072	.28	.048	.398	.398
17. DRT	.604	.178	.191	.071	.071	.155	.138	.168	.445	.486
18. FD	.268	.168	.391	.044	.279	.079	.229	.114	.304	.318
19. PC	.674	-.151	.191	.178	.078	.142	.179	.079	.609	.564

* Decimal points have been omitted

TABLE 5
Centroid Factor Loadings, Estimated and Computed Communalities of Variables for the
H₀ Group

Variable	I	II	III	IV	V	VI	VII	h _e	h _c
1. BDN	.213*	-.191	-.216	-.284	.297	.176	.146	.41	.426
2. BDP	.284	.295	.271	.115	.264	.118	.11	.39	.393
3. SO I	.585	-.162	.297	-.262	.082	.045	.261	.51	.586
4. SO II	.631	-.094	.168	-.161	.277	.076	-.127	.51	.618
5. R Comp	.651	.294	.240	.211	.211	.255	.129	.64	.688
6. DR-TR	.791	.172	.279	.074	.077	.167	.172	.64	.723
7. MP	.608	.042	-.121	.262	.220	.047	.051	.38	.396
8. IC I	.608	-.196	.129	.077	.219	.161	.128	.57	.591
9. IC II	.746	.079	.055	.046	.211	.158	.049	.65	.644
10. GI	.647	.246	.371	.076	.219	-.217	.064	.71	.774
11. Math A	.512	-.450	-.281	.265	.075	.142	-.247	.66	.722
12. Math B	.610	-.410	-.241	.167	.021	.139	.114	.75	.775
13. RP	.365	.417	.269	.248	.174	.062	-.188	.44	.482
14. 2HC	.607	.364	.152	.071	.165	.048	-.104	.45	.463
15. CC	.596	.198	.291	.129	.112	.073	.078	.51	.526
16. RC	.317	.318	-.258	.081	.168	.101	.152	.32	.350
17. DRT	.602	-.092	.184	-.067	.148	-.145	.076	.44	.456
18. FD	.137	.116	.424	.169	.127	-.191	.073	.26	.281
19. PC	.615	.122	-.071	-.325	-.178	.161	.073	.55	.568

* Decimal points have been omitted.

matrix. The stability of these r 's was not so great as for the other group, having been based on a smaller N . It might be anticipated, then, that not so many factors could be extracted from this matrix. As a matter of fact, it did not seem possible

to rotate the last factor into meaningful configuration and it was retained as a residual.

The centroid method provides orthogonal reference axes which will satisfactorily account for the relationships observed among variables. However, these centroid axes lack psychological meaning, and before an attempt is made to interpret them it is necessary to rotate the axes into more meaningful positions. To locate the new axes two principles of rotation are employed. These are: (1) that the factor loadings should be positive within the limits of sampling errors (this criterion

TABLE 6
Rotated Factor Loadings and Communalities for the Experimental Group

Variable	I	II	III	IV	Factor V	VI	VII	VIII	h ² _c	h ² _v
1. BDN	-.025*	.250	.094	-.035	.316	-.115	-.048	.120	.208	.203
2. BDP	.001	-.005	.120	-.005	.536	.185	.329	.104	.461	.455
3. SO I	.156	.061	.620	.189	.094	.025	.079	.090	.445	.459
4. SO II	.180	.080	.570	.015	.100	.335	.160	.024	.511	.512
5. R. Comp	.621	.225	.190	.028	.095	.195	-.008	.016	.521	.520
6. DR-TR	.360	.346	.545	.160	.035	-.005	.050	.207	.616	.618
7. MP	.319	.181	.081	.145	.000	.480	.451	.110	.610	.608
8. IC I	.421	.045	.555	.089	-.010	.116	.145	.210	.546	.558
9. IC II	.315	.090	.359	.158	.100	.425	.119	.225	.517	.516
10. GI	.470	-.036	.194	.040	.365	.366	.255	.005	.592	.594
11. Math A	.465	.470	.200	-.011	.075	.014	.030	.020	.482	.484
12. Math B	.494	.554	.184	-.043	-.080	.106	.120	.115	.550	.551
13. RP	.039	-.010	.100	.524	.095	.103	.165	.099	.341	.343
14. 2HC	.111	.060	.195	.417	.080	.058	.542	.100	.547	.541
15. CC	.140	.008	.300	.435	.046	.159	.355	.319	.551	.554
16. RC	-.060	.010	.040	.363	.254	.346	.262	-.010	.398	.390
17. DRT	.215	.181	.385	.205	.025	.230	.004	.365	.486	.484
18. FD	.072	-.027	.114	.440	.150	-.131	-.004	.284	.338	.333
19. PC	.175	.215	.275	.315	.327	.422	.135	.030	.564	.556

* Decimal points have been omitted.

is known as "positive manifold"); and (2) that there should be as many zero loadings in the final factor matrix as the configurations of points will permit (simple structure).

The centroid factor matrices in Tables 4 and 5 were rotated with the principles of positive manifold and simple structure in mind. The rotated factor loadings are presented in Table 6 and Table 7. Rotation was done graphically according to a method described by Zimmerman (29). Thirty-six rotations of pairs of axes were necessary to satisfy the criteria of simple structure and positive manifold for the factors extracted from the Experimental group matrix. Twenty-two rotations suf-

ficed for the factors resulting from the WASP group matrix. Loadings of less than .20 were considered as vanishing entries (i.e., approaching zero).

In Table 8 are presented the rotated factor loadings which were obtained for the group of trainees restricted on a basis of a qualifying examination. The battery is the same as the battery analyzed in the present study, with some tests having been added. This table is Table 28.18 in (12). The table is reproduced in its entirety but the order of variables and fac-

TABLE 7
Rotated Factor Loadings and Communalities for the WASP Group

Variable	I	II	III	IV	Factor V	VI	VII	h _c	h _r
1. BDN...	-.158*	.266	.169	-.049	.458	.195	-.241	.426	.421
2. BDP	-.012	.059	.054	.059	.669	.638	.679	.393	.390
3. SO I.	.110	.143	.729	.075	.134	.147	.035	.586	.577
4. SO II	.342	.115	.571	.019	.139	.179	-.059	.618	.610
5. R. Comp	.612	.467	.172	.031	.669	.689	-.264	.688	.676
6. DR-TR.	.148	.455	.494	.144	.615	.462	.059	.723	.710
7. MP	.489	.309	.129	.159	.668	.646	.149	.396	.384
8. IC I	.336	.283	.488	.625	.665	.374	.683	.591	.583
9. IC II	.482	.258	.495	.179	.122	.319	.161	.644	.630
10. GI	.742	.168	.199	.015	.479	.678	.029	.774	.765
11. Math A	.169	.787	.223	.005	.655	.115	.119	.722	.719
12. Math B	.291	.785	.168	.008	-.668	.269	.014	.776	.754
13. RP	.189	.669	.119	.648	.694	-.629	.079	.482	.474
14. 2HC.	.179	.649	.221	.532	.279	.121	.659	.463	.452
15. CC	.269	.169	.298	.489	.119	.268	-.169	.626	.609
16. RC	.189	.052	-.111	.173	.473	.189	.685	.359	.343
17. DRT	.158	.243	.499	.195	.185	.248	-.249	.456	.449
18. FD	-.008	-.104	.169	.359	-.127	.149	.271	.281	.268
19. PC	.353	.059	.469	.629	.469	.278	.175	.568	.563

* Decimal points have been omitted.

tors has been rearranged to correspond to the identifying numbers for comparable tests and factors in the present study.

Interpretation of Factors

Subjecting a matrix of r 's to a factor analysis merely indicates the number of orthogonal reference vectors which can be postulated to account for the correlations in that table. Rotating the centroid reference factors to simple structure and positive manifold tends to make for a unique solution and to make the results more meaningful psychologically. There remains the task of interpreting the factors. It has been indicated that factors simply show the presence of common refer-

ence variables for the tests analyzed. But the causes of factors are many and it is hoped, in most analyses, that the factors appearing in the particular study will indicate some of the more fundamental dimensions which may be postulated to explain the results obtained.

To determine the common causes underlying the factors obtained the usual procedure is to try to find that function which is common to all of the tests that are heavily loaded on any

TABLE 8
Related Factor Loadings and Communalities for a Group of Pilot Trainees Selected by a Qualifying Examination

Variable	Factor										Res	h ²
	I	II	III	IV	V	VI	VII	VIII	IX			
1. BDN	18*	-.03	.02	.06	.50	-.08	-.01	.05	.05	-.01	.30	
2. BDP	-.12	-.26	.10	-.03	.31	-.08	.53	.22	-.15	.03	.55	
3. SO I03	.18	.62	.03	.23	.07	-.09	.17	.00	.00	.51	
4. SO II02	.13	.56	-.08	.11	.18	.11	.24	.20	.12	.51	
5. R. Comp.58	.48	.09	-.02	.05	.04	.33	-.02	.10	.09	.70	
6. DR-TR23	.46	.38	.06	.25	.02	-.14	.36	-.02	.09	.64	
7. MP03	.20	.01	.03	-.12	.17	.58	.38	.03	.06	.57	
8. IC I29	.30	.27	-.05	.15	-.02	-.03	.40	-.10	.05	.45	
9. IC II14	.16	.28	.01	.03	.09	.22	.50	.02	.22	.48	
10. GI43	-.14	.29	-.08	-.08	-.04	.53	.23	-.06	.07	.65	
11. Math A30	.56	-.02	.02	.36	.05	.18	.08	.16	.07	.61	
12. Math B36	.58	-.08	-.04	.12	.03	.12	.26	.14	.09	.60	
13. RP	-.05	-.10	.16	.42	.03	.27	.12	.23	-.07	.21	.40	
14. zHC09	-.03	.15	.36	-.06	.10	.30	.44	-.01	-.16	.47	
15. CC01	.11	.22	.47	.04	.14	.15	.46	.01	-.17	.57	
16. RC09	.23	-.03	.10	.01	.51	.01	.28	-.16	-.05	.44	
17. DRT10	.19	.24	.24	.10	-.05	.02	.39	.07	-.03	.33	
18. FD10	.02	.22	.51	.03	-.05	-.05	.15	-.08	.02	.35	
19. PC	—	—	—	—	—	—	—	—	—	—	—	
20. History63	.08	.08	-.02	.08	-.10	.08	-.02	.52	-.05	.70	
21. Geography61	.16	.15	-.08	.01	.01	.12	.06	.58	.06	.79	
22. Physics38	.38	-.03	.04	-.11	.11	.51	.17	.11	.00	.61	
23. Decoding14	.46	.31	.12	-.06	-.15	-.17	.32	.10	.20	.55	
24. Vocabulary71	.16	.11	.10	.04	.09	-.11	.04	-.01	.11	.59	

* Decimal points have been omitted.

particular factor. Sometimes the nature of the underlying cause seems quite obvious and the factor can be quite easily identified. In other instances the nature of this underlying cause is not immediately apparent and the factor cannot be named easily. Usually some hypothesis can be formulated as to the nature of the factor and clues may be obtained which lead to further investigation. In any event, the naming and interpreting of the factors cannot be done by the statistical computations themselves, but these interpretations represent

the investigator's hypothesis as to the nature of the process which seems common to the group of variables identified by that factor.

In the present study loadings of .30 and higher are considered significant. A loading of .30 accounts for 9 per cent of the variance of a test and it would seem that this much variance should not be overlooked. In the naming of factors, however, more weight is given to those tests having the highest factor loadings. In practice, if one were attempting to purify a test so that it would be a measure of but one factor, variances even smaller than .09 would be taken into account. For the present we are concerned with identifying the factors extracted.

Factor I.—The tests having the highest loadings on factor I for all groups are (5) *Reading Comprehension* and (10) *General Information*. This factor seems verbal in nature. The mathematics tests (11 and 12) also appear to some extent. One hypothesis to account for this might be that the factor represents ability to deal with any symbolic system—not verbal symbols alone. *Mathematics B*, it will be recalled, presented items in “story problem” form, and its verbal content might be explained on these grounds. However, for the Experimental group, *Mathematics A*, which dealt primarily with algebraic equations, has a higher loading on this factor than does *Mathematics B*. Other tests also appear with minor loadings on this factor (*viz.*, (8) *Instrument Comprehension I*, (6) *Dial and Table Reading*, (7) *Mechanical Principles*, (9) *Instrument Comprehension II*, and (4) *Spatial Orientation II*). *Mechanical Principles* appears with a significant loading for the WASP group. No mechanical factor appeared for them, and one hypothesis is that this test is a generally intellective one for them. The remaining tests, it should be noted, have rather long and complicated directions in common. It is possible that comprehension of these instructions (verbal comprehension) is represented here. The general nature of the function represented by the factor seems quite clear. It has been designated verbal (V).

Factor II.—Tests having consistently high loadings for this factor for each sample are: (12) *Mathematics B*, (11) *Mathematics A*, and (6) the *Dial and Table Reading Test*. An inspection of the tests to determine the common element

immediately suggests the numerical content. This factor, accordingly, can be designated as numerical (N). The relatively high loading of the *Reading Comprehension Test* on this factor for the WASP and Restricted groups is of interest, as are the loadings of the mathematics tests on the verbal factor. It is possible that there is an underlying second-order factor which might be described as a "general intellective" factor reflected in these loadings.

Factor III.—The *Spatial Orientation Tests* (3 and 4) have consistently high loadings on factor III for all groups. The loadings of the *Instrument Comprehension Tests* (8 and 9) and *Dial and Table Reading Test* (6) are somewhat lower. These tests have in common the requirement that the subject be able to pick out and react to details in a rather complicated presentation of stimuli. In the *Spatial Orientation I Test* the subject has to match small parts of a photograph with the corresponding area in a large photograph. In *Spatial Orientation II Test* small photographs are matched with corresponding areas on maps. In *Instrument Comprehension I* relevant aspects from a set of dials must be picked out from a group of five items describing the dials. In the *Dial and Table Reading Test* dials and tables must be read quickly and accurately. The speed with which one can accomplish each of these tasks is associated with the score made on the test. This factor has been called the perceptual speed factor (P).

Factor IV.—The four most consistent tests for this factor are *Rotary Pursuit* (13), *Finger Dexterity* (18), *Complex Coordination* (15), and the *Two-hand Coordination Test* (14). The common element is immediately apparent—they are all psychomotor tests. All the tasks involve some form of motor response and of coordination between various members of the body. The *Rotary Pursuit Test* involves eye-hand coordination and rhythmic use of the hand and arm. *Finger Dexterity* involves the use of smaller muscle groups and a close coordination between these and the eyes. The *Complex Coordination Test* involves eye, hand, and foot coordination of a relatively gross nature. The *Two-hand Coordination Test* involves the simultaneous use of both hands and the coordination of these movements with visual cues. This factor has been called the psychomotor coordination factor (PM).

Factor V.—An examination of factor loadings shows factor V is comparable only for the Experimental and WASP groups. Tests having significant projections are: *Biographical Data Blank—Pilot* (2), *General Information* (10), and *Biographical Data Blank-Navigator* (1). The variable common to these tests seems to be an interest in aviation. Both biographical data blanks contain numerous items which indicate an interest in aviation, as does the *General Information Test*. The individual who read articles about the various planes would do well on the first part of this test, while the individual who familiarized himself with technical terms and those specific to flying would do well on the second part. It seems a reasonable hypothesis that this factor reflects interest in flying, and it has been labeled the aviation interest factor (AI).

No factor called aviation interest was isolated for the Restricted group. Factor V for this group was called "mathematical background" apparently because of its projections on the *Biographical Data Blank—Navigator* and on *Mathematics A*. The validity of mathematics for navigators had been demonstrated. However, a minor projection appears on this factor for *Biographical Data Blank—Pilot*, and this is rather difficult to interpret in terms of mathematics background. Also unusual, if this is mathematics background, would be the lack of a significant loading for the other mathematics test. It is probable that further rotation might make it more meaningful. At best the factor has but one significant projection for this group—that on *Biographical Data Blank—Navigator*, and it is difficult to interpret on this basis alone.

Factor VI.—Factor VI contains no very significant loadings for any of the groups. For the Experimental and WASP groups it seems most closely related to a factor which has been called visualization (Vz) in other analyses. This is suggested by the loadings on the *Instrument Comprehension Tests* and on the *Complex Coordination Test*. However, the loading for the *Dial and Table Reading Test* for the WASP's cannot be accounted for readily on this basis. The interpretation of this factor must remain rather tentative. For the Restricted group Factor VI was called a kinesthetic factor, but again this is rather doubtful since it has but one significant loading (in Rudder Control).

Factor VIII.—For the WASP group this factor could not be rotated into meaningful position and has been called residual. For the men's populations, however, it has loadings in *Mechanical Principles*, *Two-hand Coordination*, and, to some extent, in *Biographical Data Blank—Pilot*. All of these have already shown major or minor projections on one of the factors previously discussed. An examination of these tests suggests that the common element is mechanical experience. It seems reasonable that mechanical experience might be associated with performance on the *Two-hand Coordination Test*, where the handles with which the stylus is manipulated resemble the handles on a lathe. The mechanical content of the *Mechanical Principles Test* is obvious. The *Biographical Data Blank—Pilot* also contains items which indicate mechanical experience. This factor has been called mechanical experience (ME).

Factors VIII and IX.—For the Restricted group Factor VIII was called space (SP). For the Experimental group there are no very significant loadings but the trend is such that its similarity to this factor is suggested. Factor IX, found only in the Restricted group, was called a social science background factor.

General interpretation.—It has been possible to identify five of the factors found in the analyses of the Experimental and WASP groups with a fair degree of assurance. These factors were found in parallel form in the Restricted group. The factors which could be identified were verbal, numerical, perceptual speed, psychomotor coordination, and aviation interest. It was not so easily possible to interpret the remaining factors. One seemed clearly to be mechanical experience. A tentative hypothesis is that the others represent abilities called visualization and space. Because of the dearth of significant loadings, however, their identifications must remain rather tentative. If tests that were purer measures of these factors had been included in the battery it might be anticipated that the nature of these factors might be made more clear.

Summary and Conclusions

The purpose of this study was to determine empirically the effects of selection of samples upon factors isolated from a battery of tests. Three samples of pilot trainees representing dif-

ferent selective conditions had taken a comparable battery of tests. Correlation matrices were factorially analyzed for each group by the centroid method. The resulting factors were compared. It was possible to conclude that similar primary factors will be obtained when a battery of tests is administered to groups representing different selective conditions. The results suggest that primary factors are meaningful descriptive parameters which are not merely reflecting the particular selective conditions represented in the various groups.

REFERENCES

1. Balinsky, B. "Analysis of the Mental Factors of Various Age Groups from Nine to Sixty." *Genetic Psychology Monographs*, XXIII (1941), 190-234.
2. Blackwell, A. M. "A Comparative Investigation into the Factors Involved in Mathematical Ability of Boys and Girls." *British Journal of Educational Psychology*, X (1940), 146-153, 212-222.
3. Buxton, C. F. "The Application of Multiple Factorial Methods to the Study of Motor Abilities." *Psychometrika*, III (1938), 85-93.
4. Clark, M. P. "Changes in Primary Mental Abilities with Age." *Archives of Psychology*, XLI (No. 291, 1944), 1-30.
5. Coombs, C. H. "A Factorial Study of Number Ability." *Psychometrika*, VI (1941), 166-180.
6. DuBois, P. H. (ed.) *The Classification Program*. Report No. 2 of the AAF Aviation Psychology Program Research Reports. Washington, D. C.: Government Printing Office, 1947.
7. Flanagan, J. C. "The Experimental Evaluation of a Selective Procedure." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, VI (1946), 445-466.
8. Garrett, H. E., Bryan, A. I. and Perl, R. E. "The Age Factor in Mental Organization." *Archives of Psychology*, XXVI (No. 176, 1935), 1-31.
9. Goodman, C. H. "Factorial Analysis of Thurstone's Primary Abilities." *Psychometrika*, VIII (1943), 121-129.
10. Guilford, J. P. *Psychometric Methods*. New York: McGraw-Hill Book Co., Inc., 1936.
11. Guilford, J. P. "The Difficulty of a Test and its Factor Composition." *Psychometrika*, VI (1941), 67-77.
12. Guilford, J. P. (ed.) *Printed Classification Tests*. Report No. 5 of the AAF Aviation Psychology Program Research Reports. Washington, D. C.: Government Printing Office, 1947.
13. Melton, A. W. (ed.) *Apparatus Tests*. Report No. 4 of the AAF Aviation Psychology Program Research Reports. Washington, D. C.: Government Printing Office, 1947.

14. Richards, T. W. "Genetic Emergence of Factor Specificity." *Psychometrika*, VI (1941), 37-42.
15. Schiller, B. "Verbal, Numerical, and Spatial Abilities of Young Children." *Archives of Psychology*, XXIV (No. 161, 1934), 1-69.
16. Thomson, G. H. "The Influence of Univariante Selection on the Factorial Analysis of Ability." *British Journal of Psychology*, XXVIII (1938), 451-459.
17. Thomson, G. H., and Ledermann, W. "The Influence of Multivariate Selection on the Factorial Analysis of Ability." *British Journal of Psychology*, XXIX (1939), 288-306.
18. Thurstone, L. L. *Vectors of Mind*. Chicago: University of Chicago Press, 1935.
19. Thurstone, L. L. "Current Misuse of the Factorial Methods." *Psychometrika*, II (1937), 73-76.
20. Thurstone, L. L. "Primary Mental Abilities." *Psychometric Monographs*, I. Chicago: University of Chicago Press, 1938.
21. Thurstone, L. L. "The Effects of Selection in Factor Analysis." *Psychometrika*, X (1945), 165-198.
22. Thurstone, Thelma G. "Primary Mental Abilities of Children." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, I (1941), 105-116.
23. Wittenborn, J. R. "Mechanical Ability, Its Nature and Measurement. I. An Analysis of the Variables Employed in the Preliminary Minnesota Experiment." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, V (1945), 241-260.
24. Wittenborn, J. R. "Mechanical Ability, Its Nature and Measurement, II. Manual Dexterity." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, V (1945), 395-409.
25. Wolfe, D. "Factor Analysis to 1940." *Psychometric Monographs*, III. Chicago: University of Chicago Press, 1940.
26. Woodrow, H. "The Common Factors in Fifty-two Mental Tests." *Psychometrika*, IV (1939), 99-108.
27. Zimmerman, W. S. "A Simple Graphical Method for Orthogonal Rotation of Axes." *Psychometrika*, XI (1946), 51-55.
28. Staff, Field Research Unit, AETRC, and MPEU #8, Sheppard Field. "Performance of WASP's in Training Upon Aviation Cadet Qualifying and Classification Tests." *Research Bulletin* Hq. 44-25, (Published by the Psychological section, Office of the Surgeon, Hq., AAFTC), May, 1944.

THE ROLE OF CLIENT TALK IN THE COUNSELING INTERVIEW

EARL F. CARNES AND FRANCIS P. ROBINSON

Ohio State University

RECENTLY a number of studies have been made of the dynamic relationships involved in the counseling interview (3, 4, 8, 9). A promising methodology for studying the interview has evolved; a recent trend has been an examination of the relationship of single variables within the interview to other counseling procedures and outcomes (1, 10). The present study is an example of that trend.

Extensive counselee talk is considered by many to be an important attribute of effective counseling (2, 3, 6). In the first place, a counselee's willingness to talk is usually symptomatic of a good working relationship between client and counselor. A reticent or resistive client tends to speak briefly or even to be silent; free talking is taken as a good sign. As a result, many counselors try consciously or unconsciously to enhance its occurrence. Extensive counselee talk is also valued by many for a second reason. It is felt that free verbalization tends to bring more material out at a conscious level and until there is verbalization by the counselee, there can be little growth in insight. The assumption, apparently, is that verbalization results in insight. Like most other variables in the counseling area, however, the importance of client talk has not been objectively studied. This paper proposes to examine the relationship of the amount of counselee talk to the effectiveness of counseling.

Method

Source of Data.—The interview typescripts used as the raw data for the study were obtained from a counseling practicum offered at Ohio State University. About half of the data came from interviews conducted by experienced counselors, some of

whom were supervisors of the course and others of whom were not ordinarily a part of the course. The remainder of the data came from the counselors-in-training. Counselees consisted of students in a how-to-study course, who met for a weekly interview with counselors over a period of a quarter to discuss problems related to their study difficulties (5). While study problems are emphasized, clients bring up all types of problems.

The interviews were recorded and verbatim typescripts made of the recordings. Use was made of 78 such typescripts in this study. Of this number 42 interviews were conducted by 7 experienced counselors and the other 36 by 16 counselors-in-training.

Classification System.—In the analysis of an interview, the most advantageous unit has been found to be all the counselor and counselee discussion related to one main idea or problem. Sherman (8) found it possible to recognize with a high degree of consistency the transition points at which the conversation changes from one topic to another, and the unit then consists of all the conversation between two transition points. A typical interview may consist of one or many such units. For the 78 interviews used in this study, the range was from 1 to 15 units per interview. The total number of units used was 353.

The most meaningful measure of the amount of client talk is a ratio of the amount of client talk to the total amount of talk in the interview. For ease of reference this ratio will be called the "talk ratio" and will be presented in decimal fractions. As an example, if the talk ratio is .25, it is apparent that the counselee has done 25 per cent of the talking.

It is proposed to study the relationship of the amount of client talk to the topics, techniques, and outcomes of interview discussion units. While some other topics were discussed, this study will be made only on the units dealing with the four most common topics, i.e., study skills, vocational problems, therapy, and scholastic questions (information seeking).

In any given unit several techniques may be commonly used, but it has been found that one usually stands out as predominant in the conduct of the interview. This has been labeled the "primary technique." Each counselor response in a unit was tabulated on a check list of the various possible techniques and

the primary technique was then determined on the basis of its frequency or importance for the total unit. In the present study, units were included which were characterized by the primary use of one of four techniques. The definitions of these four, arranged in the order of increasing amounts of counselor lead, follow:

1. *Clarification*: The counselor states the feeling or problem of the counselee in such a way that it can be more clearly understood and used in the interview, but he does not interpret the counselee's feelings or problem, nor does he add any information or suggestion of his own.
2. *Tentative Analysis*: The counselor presents a conception of the problem being discussed but does so in a purely tentative manner. The client feels freedom to accept, modify, or reject the idea. The counselor is not attempting to push the counselee toward a given idea, but is setting up ideas for analysis. If, following a statement of the counselee, the counselor asks a question of the "tell me more about that" type, it is also considered tentative analysis.
3. *Interpretation*: The counselor verbalizes his conception of the problem and states it positively enough that the counselee feels that it is presented for his acceptance. Data may or may not be given to support the interpretation. As presented the interpretations seem objective and obvious and the counselor's values do not seem to be involved.
4. *Urging*: The counselor not only presents his conception of the problem but gives advice on its treatment. The client definitely feels that he is being pushed toward acceptance. It often contains phrases like "I want you to" or "You should." Data may be used to support the argument. The counselor's values seem to be involved here.

As evidences of interview outcomes, three variables were used: (1) growth in counselee insight, (2) working relationship, and (3) client responsibility for the progress of the interview. These factors are admittedly difficult to objectify, but the most feasible means of doing so seemed to be by the use of rating scales. Five-point rating scales, carefully worked out by Sherman (8), were used in this study. While the reliability of results obtained from rating scales are not particularly high, Sherman found that these rating scales produce results adequate for the comparison of group averages used here.

A distribution of the units used in the study, classified by topic and primary technique, is shown in Table 1.

Results

An analysis of the 353 units shows that there was a wide range in the amount of counselee talk from one unit to another. One counselee talked as little as 4 per cent of the total, another talked 89 per cent of the total. The individual counselors differ considerably from one another in the mean talk ratio of their units, but each tends to be consistent from one unit to another. Among the nine counselors who contributed the largest number of units to the study, the individual mean talk ratios ranged from a low of .26 with a standard deviation of .15, to a high of .67 with a standard deviation of .21. The data were analyzed to determine whether the experience of the counselor had any effect on the talk ratio, but the experienced

TABLE 1
Distribution of Units Classified by Topic and Primary Technique

	Clarification	Tentative Analysis	Interpretation	Urging	Totals
Study Skills	24	22	65	36	147
Scholastic Questions	33	23	37	7	100
Therapy	13	11	33	9	66
Vocational Problems	1	10	22	7	40
Totals	71	66	157	59	353

counselors and the counselors-in-training, as groups, had almost identical distributions of talk ratios.

The relationship between the talk ratios and the rated amount of insight growth achieved by the counselee on each unit was studied. The correlation between these two variables for all units is $.25 \pm .05$. Sherman (8) found that insight ratings were her least reliable measure. Judges had particular difficulty in rating short responses. What is one to infer about insight from a counselee response of "Mm hm"? It is quite possible that judges tend to rate down brief counselee remarks, in which case a low talk ratio would tend to receive a low insight rating. If we accept Shaffer's (7) definition of insight as "a verbalization of the person's problems and of the relationships between them," perhaps such a low rating is justified.

Logically the amount of counselee talk seems to be a function of the working relationship. In the low ratings of working relationship, where the counselee is resistive, he is apt to speak

in monosyllables or perhaps to merely nod his head in silence. On the other hand in the high ratings, there is mutual give and take and each party feels free to talk. The coefficient of correlation is $.31 \pm .05$. Further details of the relationship between the two variables is evident from a glance at Table 2. The distribution shows that working relationship is virtually always judged to be good when the talk ratio is high; out of 71 units with a talk ratio over .60, only 8 show a working relationship rating of less than 4. However, a high talk ratio is not the *sine qua non* of a good working relationship; this is shown by the fact that out of 181 units with a talk ratio of less than .40, 98 were still classed in the highest two ratings of working relationship. In short, a high talk ratio is not necessary to a good working relationship, but it is good insurance.

TABLE 2
Distribution of Talk Ratios of Units According to Their Working Relationship Scores

Working Relationship	0-.19	20-.39	Talk Ratios		.80-.99	Total
			40-59	60-79		
5	3	24	19	20	8	74
4	9	62	62	31	4	168
3	30	27	8	1	0	66
2	5	17	12	5	1	40
1	1	3	0	1	0	5
	48	133	101	58	13	353

The most clear-cut relationship found in the study is between the talk ratio and counselee responsibility for the progress of the unit. The coefficient of correlation is $.66 \pm .03$. It is evident that the person feeling primary responsibility for directing the course of the interview tends to talk the most.

Up to this point, overall relationships have been examined. The question now arises, what happens to these relationships when different topics or different primary techniques are studied? It may be that the amount of talk is more important with some problems than others. One would logically expect, for example, that insight concerning study skills would come from having an expert explain effective study techniques more than from counselee verbalization of his difficulties, since effective study depends upon information as to correct techniques. In therapy, on the other hand, there is more need for

growth in self-understanding through verbalization of vague fears than for the acquisition of information.

When the units are grouped by topics and the talk ratios are tabulated, it turns out as expected that the lowest mean ratio (.36) occurs in the units concerned with study skills. The other three topics, in which the counselee discusses his problems, all have about the same mean talk ratio (.43-.48). Therapy, with its emotional loading, does not here result in more talk than other personal problems.

An analysis was made to determine whether the relationships between talk ratio and growth in counselee insight and between talk ratio and working relationship were affected by the topics of the units. Results are shown in Table 3.

The topic of the unit apparently has a marked effect on these relationships. An explanation for the low correlation between

TABLE 3
Analysis by Topics of the Correlation of Talk Ratios to Insight and to Working Relationship

Topic	Growth in Insight vs. Talk Ratio (r)	Working Relationship vs. Talk Ratio (r)
Study Skills.09	.35
Scholastic Questions.43	.49
Therapy.25	.11
Vocational Problems.60	.61

talk ratio and insight in study skills has previously been suggested, i.e., growth in insight in study skills may not necessarily come from client conversation so much as from counselor description of correct study techniques. However, the low correlation in therapy looks out of line. Because the sampling of counselors in this study included some persons highly interested in a non-directive approach to therapy problems, the counselors who conducted the therapy units were divided into two groups: those considered essentially directive and those essentially non-directive. This grouping was made without knowledge of the talk ratios by a staff member, who has been closely associated with all of the counselors involved. The distributions of talk ratios for the directive and the non-directive counselors are shown in Chart I.

On the basis of this division, the 23 therapy units conducted

by non-directive counselors yielded a mean talk ratio of .66 and an average insight score of 2.35. The 43 directive units yielded an average talk ratio of only .29, but also an average insight score of 2.42.

Coefficients of correlation between insight and talk ratio were computed for each of these two groups. For the non-directive group, the coefficient was $.39 \pm .18$, and for the directive group it was $.36 \pm .12$, as compared to $.25 \pm .05$ for the two groups combined. The reason for this is that in non-directive therapy there is a bunching of units at the high end of the talk-ratio scale, whereas in the directive therapy there is a bunching

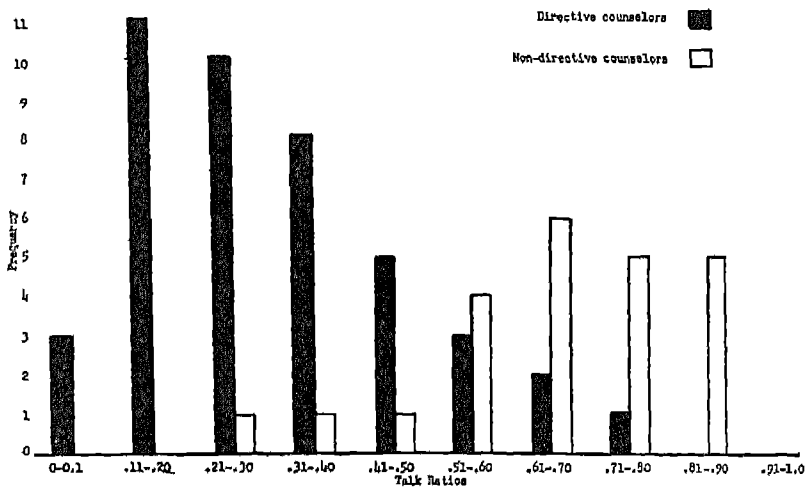


CHART 1. Distribution of Talk Ratios of Directive and Non-directive Counselors

at the low end of the scale. Combining the two groups causes a large standard deviation of talk ratio, which in turn decreases the correlation coefficient.

These results for directive and non-directive counselors show that the amount of client talk is related to insight, and that clients of non-directive counselors talk the most. But the results also show that it is the relative amount of talk for a given counseling system, and not the absolute amount of talk that is related. The best probable interpretation of these results is that the system of counseling tends to set limits to the amount of counselee talk, and within these limits counseling

situations giving rise to insight lead to more talking. It is not clear that trying to get more counselee talk will necessarily lead to more insight.

A similar analysis of the low correlation in therapy units between talk ratio and working relationship yields a somewhat similar result in that both the directive and the non-directive groups alone get higher correlations than do the two combined. However, in this case the correlation coefficient for non-directive therapy is considerably higher ($.38 \pm .18$) than in directive therapy ($.17 \pm .15$).

The final factor in this investigation which appears to be related to the talk ratio is the primary counselor technique. The distribution is shown in Table 4.

The primary techniques are arranged in order of increasing amounts of counselor lead. It is apparent that in general there

TABLE 4
Relationship Between Talk Ratio and Primary Technique

	Talk Ratio					Mn	SD
	0-.19	.20-.39	.40-.59	.60-.79	.80-.99		
Clarification	3	14	32	16	6	.51	.19
Tentative Analysis	0	19	24	19	4	.52	.18
Interpretation	21	71	41	21	3	.39	.19
Urging	24	29	4	2	0	.26	.15

is an inverse relationship between the amount of lead and the talk ratio. Little difference is seen, however, between the talk ratios for clarification and tentative analysis.

The relationship between amount of lead and talk ratio suggests that the amount of lead has influenced the correlations between talk ratio and interview outcomes. In a similar way the topic has exerted an influence. A rough control of these two influences may be obtained by using derived scores, i.e., by grouping units according to topic, subgrouping the topics by primary technique and subtracting from every member of each subgroup a figure equal to the mean talk ratio for the subgroup. Thus a client who talks more than the average client does when a particular topic is discussed and a particular counseling technique is emphasized, probably wants to talk. Is this measure of willingness or lack of willingness to talk related to interview outcomes? Such a method should give a somewhat

more accurate picture of the relationship between talk ratio and outcomes, independent of techniques and topic

For the topic of study skills, the correlation between talk ratio and insight, using derived scores, is $+.04$, and the correlation between talk ratio and working relationship is $+.12$. For the decision-making topics, the correlation between talk ratio and insight is $+.39$, and between talk ratio and working relationship is $+.23$. There is again an indication that in study skills, talk ratio is not an important factor. It seems to be somewhat more important in the decision-making units.

While this study has consistently shown low positive relationships between amount of client talk and worthwhile outcomes, the causal relationship is not clear. Detailed analysis has shown that counseling systems tend to set patterns affecting the amount of talk, within these limits, the relationship of talk and outcomes is higher. Insofar as the amount of talk is a symptom of outcomes it may be welcomed as a sign of effective work by the counselor. Insofar as increased amount of talk may aid in clarifying problems, the counselor may build his confidence to obtain it. While the direction of causality and the degree of importance of the possibilities cannot be assessed, it is clear from the evidence at hand that very low ratios of counselee talk are often bad. Great care will have to be taken, however, not to overemphasize amount of talk as a criterion in counseling. Emphasis on the amount of counselee speech may direct the conference to safe topics on which the client talks easily or even to social visiting. Amount of talk should be considered only a very minor criterion of counseling effectiveness.

Summary and Conclusions

Typescripts of 78 counseling interviews (353 discussion units) were analyzed to determine the relationship between the amount of client talk and the following variables: topic, counselor technique, growth in counselee insight, working relationship, and counselee responsibility for the progress of the unit. Amount of client talk yielded a low positive correlation with growth in client insight and with working relationship, and a more marked correlation with client responsibility for the progress of the interview.

The topic of the unit exerts an influence on the relationship between amount of client talk and insight growth. In study skills this relationship is virtually zero. In various types of personal problems, it is large enough to seem important. In therapy, non-directive counselors obtained much more counselee talk than did directive counselors but these large amounts of talk were no more highly related to insight. The topic of the unit exerts a similar influence on the relationship between client talk and working relationship.

In general there is an inverse relationship between the amount of lead in a given counselor technique and the amount of client talk.

The causal relationship between amount of client talk and desirable interview outcomes is not clear. Therefore, it is not possible to use the amount of client talk as a criterion of counseling effectiveness.

REFERENCES

1. Daulton, M. J. "A Study of Factors Relating to Resistance in the Interview." Unpublished M.A. Dissertation, Ohio State University, 1947.
2. Hobbs, N. "Psychological Research and Services in an Army Air Forces Convalescent Hospital." Unpublished Ph.D. Dissertation, Ohio State University, 1946.
3. Porter, Jr., F. H. "The Development and Evaluation of a Measure of Counseling Interview Procedures." *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, III (1943), 105-126, 215-238.
4. Raimy, V. C. "The Self Concept as a Factor in the Counseling and Personality Organization." Unpublished Ph.D. Dissertation, Ohio State University, 1943.
5. Robinson, F. P. "Two Quarries with a Single Stone." *Journal of Higher Education*, XVI (1945), 201-206.
6. Rogers, C. R. *Counseling and Psychotherapy*. New York: Houghton-Mifflin Co., 1942.
7. Shaffer, L. F. "The Problem of Psychotherapy." *American Psychologist*, II (1947), 459-467.
8. Sherman, D. M. "An Analysis of the Dynamic Relationships Between Counselor Techniques and Outcomes in Larger Units of the Interview Situation." Unpublished Ph.D. Dissertation, Ohio State University, 1945.
9. Snyder, W. U. "An Investigation of the Nature of Non-Directive Psychotherapy." *Journal of General Psychology*, III (1945), 193-224.
10. Tindall, R. H. and Robinson, F. P. "The Use of Silence as a Technique in Counseling." *Journal of Clinical Psychology*, III (1947), 136-141.

THE RELATIONSHIP BETWEEN VALUES, CONCEPTS, AND ATTITUDES

ASAHEL D. WOODRUFF

and

FRANCIS J. DIVESTA

Cornell University

PREVIOUS studies of the value patterns of individuals have shown that values are clearly related to behavior (5, 11, 13, 18, 19). The relationship seems to be something along the following lines. As a result of experience, the individual comes to value positively certain objects and conditions which have seemed to contribute to his well-being. Those objects or conditions which seem to him to have interfered with his well-being become colored with a negative value. Variations in the value attachments presumably depend on the amount of contribution or interference the individual has experienced in connection with the objects or conditions. The result of this process is a pattern of values running from high positive through neutral to high negative strength. In any behavior situation, an individual tries to act in a way which will get him more of his positive values and less of his negative values.

Before dealing with values in research studies, it is necessary to become somewhat clear as to the level of generality or specificity involved in the value categories to be used. Value can be attached at such levels of specificity or generality as a particular food, a pet, a type of social event, a general characteristic of people, or a generalized condition of living. Many of the most used concepts in American society are of the latter type and include such concepts as having home life, possessing wealth, engaging in excitement, doing socially useful things, and so on. Accordingly the senior writer has defined a "value" as a generalized condition of living which the individual feels has an important effect on his well-being. Individuals who value home life may differ as to the details which make up their

concepts of home life, but they do value home life more, for example, than excitement, or personal improvement, or something else. Hence they will conduct their affairs so as to promote more home life rather than less home life, and in preference to more excitement, or personal improvement as the case may be. Speaking generally, an individual will try to promote all of his higher positive values as opportunity permits, but each value will have precedence over those below it in his pattern, when a conflict between values develops.

Preliminary Trial

One of the interesting hypotheses which grows logically out of the belief that values are governing factors in behavior, is that values play something of an important role in the determination of expressed attitudes. If this can be established as a fact it will have an important effect on research in attitudes. One thing it will demonstrate is that attitudes are behavioral responses to something more central and more stable in the individual's mental make-up.

As a preliminary approach to this hypothesis, a group of 72 students in a first course in psychology were subjected at the beginning of the term to a series of tests. *A Study of Choices* (17, 18) was administered to obtain their value patterns, and Thurstone's scales for measuring attitude toward Communism, Sunday Observance, and Evolution were administered. For each student a data card was made up which showed his attitude scores on the three attitude scales, and the rank in his pattern of each of the twenty-four values contained in the test of values. Intercorrelations were computed between all of these variables, and the values were arranged in three lists, one for each attitude, in the order of the intercorrelations as shown in Table 1.

Inferences For Studying Subjective Views

Nothing conclusive is established by these data, but they are highly suggestive. There seems to be a clear tendency for the intercorrelations to be high positive when the subject of the attitude score would seem, on a commonsense basis, to contribute toward the value in question. Thus, Sunday obser-

TABLE 1
Correlations Between Twenty-four Values, and Attitudes Toward Communism, Sunday Observance, and Evolution

Communism		Sunday Observance		Evolution	
Value	r	Value	r	Value	r
Intellectual conversation.....	.38	Church activity. . .	.49	Exciting work. . .	.41
Civic political power.35	Religious living45	Intellectual conversation30
Intellectual work.33	Personal social service33	Pure excitement.27
Many social friends33	Family life21	Relaxation and recreation. . .	.25
Promoting social betterment . .	.30	Promoting social betterment . .	.16	Control of wealth.21
Personal social service.22	Fine home setting.13	Physical comfort.18
Security for many21	Informal leadership07	Intellectual work.14
Security for self.18	Social ease and poise07	Civic political power13
Informal leadership.17	Informal social activity05	Informal social activity10
Family life.17	Many social friends.02	Friendships in business07
Friendships in business.04	Dominating personality	-.02	Formal social activity07
Possession of wealth.	-.02	Possession of wealth.	-.04	Possession of wealth.07
Relaxation and recreation . . .	-.03	Security for many.06	Dominating personality.06
Control of wealth	-.03	Formal social activity	-.10	Many social friends02
Fine home setting.	-.03	Physical comfort.12	Security for many	-.05
Church activity05	Intellectual conversation14	Social ease and poise.10
Religious living.06	Civic political power15	Security for self13
Physical comfort	-.07	Friendships in business17	Fine home setting.18
Exciting work17	Pure excitement18	Personal social service20
Dominating personality28	Intellectual work18	Family life23
Social ease and poise32	Security for self19	Promoting social betterment . .	.23
Pure excitement37	Control of wealth22	Informal leadership.39
Formal social activity40	Exciting work.22	Church activity.41
Informal social activity52	Relaxation and recreation28	Religious living.47

vance would seem to be contributory to the realization of church activity, and religious living, and in a lesser sense to the promotion of service to others, and to family life. Also, Sunday observance would seem to be detrimental to the realization of relaxation and recreation, and to excitement. Furthermore, the values which have intercorrelations around zero seem, on the whole, not to be greatly affected by the subject of the attitude test.

In arriving at the tentative conclusions just outlined, it is necessary to reach for what one might conceive to be the subjective interpretations of the typical undergraduate student. The objective facts about any existing relationship between Sunday observance and recreation are of no importance here, because we are dealing with the subjective and personal reactions of individuals. Therefore, it becomes necessary to recognize certain stereotyped concepts such as the idea that evolution is contrary to traditional religious thinking and is therefore anti-church and even anti-religious living. Under that assumption we obtain what seem to be reasonable relationships between values and attitudes, but it is clearly necessary to be much more precise as to what each individual's concepts are if we are to trace the relationship between values and attitudes with real validity.

Revision of Hypothesis

With this hunch, the hypothesis was revised to read as follows. An individual's attitude toward any object, proposition, or circumstance will be *favorable* if, *according to his concepts*, that object seems to favor the achievement of his strong positive values. His attitude will become *unfavorable* if any change in the situation, or in his concept of the object, makes it seem to him that the object endangers his strong positive values. Conversely, one's attitude toward any object, proposition, or circumstance will be *unfavorable* if, *according to his concepts*, the object seems to threaten his strong positive values. His attitude will become *favorable* if any change in the situation or in his concept of the object makes it seem to favor his strong positive values. The strength of the attitudinal expression will be a function of the importance of the

values to which the object or condition has any relationship and the extent to which the person feels the object or condition will affect his values.

Procedure

A new class of 84 students just beginning the first course in psychology were used as subjects. *A Study of Choices* (17) was used to obtain the value pattern of each one. Remmers' *Generalized Scale for Measuring Attitude Toward Any Proposed Social Action* was used to determine the students' attitudes toward the proposed abolition of fraternities and sororities from American colleges. Each student completed a trial form of a concept test, to obtain a description of his concept of fraternities or sororities. The concept test consisted of two parts. One was a sheet on which appeared the following material, with appropriate spaces for writing:

If the abolition of sororities and fraternities were adopted by American colleges it would have certain effects on individuals such as you. Using your best judgment, even though you may not have facts upon which to base it, make a brief answer to each of the following questions.

1. What would happen to YOU?

Would it be good —, bad —, or neutral —?

2. What would happen to YOUR BEST FRIENDS?

Would it be good —, bad —, or neutral —?

3. What does it mean? What would it be like? Tell what you think the action would involve or include? In other words, describe what you think would take place.

The other sheet was set up in the form of a rating scale on which a series of questions could be answered by checking one of five positions: Very Detrimental, Mildly Detrimental, Neutral, Mildly Beneficial, and Very Beneficial. The instructions and the questions were as follows:

CHECK LIST ON SORORITIES AND FRATERNITIES

Place a check mark on the line below each question in a position which indicates the amount of benefit or detriment you believe is involved. The center of the line is always a neutral point, with beneficial effects to the right and detrimental effects to the left. Use the descriptions above question number 1 as a guide.

1. How do they affect their members socially during college years?
2. How do they affect members scholastically during college?
3. How do they affect members financially during college?
4. How do they affect members morally during college?
5. How do they affect members socially after college?
6. How do they affect members financially after college?
7. How do they affect members morally after college?
8. How do they affect non-members socially during college?
9. How do they affect non-members scholastically during college?
10. How do they affect non-members financially during college?
11. How do they affect non-members morally during college?
12. How do they affect non-members socially after college?
13. How do they affect non-members financially after college?
14. How do they affect non-members morally after college?
15. How do they affect American social life generally?

The concept tests are obviously constructed on the assumption that things mean to us what we believe they *do* to us. This is clearly substantiated by the responses of children to the Binet, and the answers which are credited at the early age levels. It is also supported by the role theory of Cottrell (2), who speaks of these functional meanings as self-other roles incorporated in the individual through his interactions with other social objects.

The test of values was scored in the usual manner, and for each student a data card was set up on which his rank-order arrangement of the twenty-four values appeared. In the scoring procedures for this test (18) it is possible to determine whether the individual discriminates between two subdivisions of a generalized value. If he did not do so, those two elements were listed as one in this study. This is significant, because only the five highest values, whether they were subdivided or not, were used in each case in subsequent treatment of the data. Thus the values used in one person's case might not be the same as those used in the case of another person, except that they would always be the highest five for each student. This was done in order to restrict the test relationship to the strongest values for each person, and thus let significant individual differences exert their maximal weight in the results.

The attitude test was scored in the usual manner, and the score recorded on the data sheet for each student.

The concept check-sheet was scored in two steps. First the five conceptual elements in it were jotted down on the side of the sheet. These were social life, intellectual life, moral life, economic life, and general American social life. Next a count of the check marks was made, in which zero checks were ignored, "1" checks were counted as 1, and "2" and "3" checks were counted as 2. Those on the beneficial side were given plus ratings, and those on the detrimental side were given minus ratings. Opposite each of the five elements was shown the algebraic total of the check marks. Thus one student had plus 2 and minus 1 for society, minus 1 for moral, and minus 1 for economic. All of his other checks were at the zero position or neutral. This summarizes his expressed concept of what a fraternity does to a person, namely, it has mildly beneficial social effects and mildly detrimental moral and economic effects, but in the main it has very little total effect. In comparison with most of the students, this person's concept is very close to being a blank, an interesting phenomenon which will be discussed later.

The other concept sheet, requiring free-response answers, was read carefully for any content which revealed the individual's functional concept of a fraternity or sorority. For example, here is one response to the first question: "What would happen to YOU?"

First, I would not have as pleasant an opportunity to work (1), as working in a fraternity has helped me considerably towards seeing me through college (2). I would have no "home" at college (3), and none of the close friendships that I enjoy (4). I would not have the pleasure of working and living with a group (5), nor would I be able to enjoy the outside friendships (6) to which the group has introduced me.

The numbers inserted in the quotation are to identify the interpretations, which were listed as follows:

- (1) Comfort, plus 1.
- (2) Economic advantage, plus 2.
- (3) Home life, plus 2.
- (4) Friendship, plus 2.
- (5) Society, plus 1.
- (6) Friendship, plus 1.

A notation of "plus 1" indicates that the scorer felt this comment indicated a belief that a fraternity contributed mildly to one's comfort. A notation of "plus 2" indicates the comment showed a belief in a somewhat stronger effect. In the quotation above, for example, the word "considerably" seemed to give emphasis to the comment. In view of the subjective nature of this scoring, both experimenters scored the papers separately, and then discussed those papers on which they differed. Such papers were in the minority. This was done before either experimenter had an opportunity to associate a value pattern or an attitude score with the name of the individual whose sheet was under discussion.

In order to equate these quantitative concept scores to the value pattern, "moral" was assumed to be roughly equivalent to the value "religion," "social" to "society" in the value test, "financial" to "wealth" in the value test, and "scholastic" to "intellectual activity" in the value test.

The scores derived from each student's concept tests were next set opposite the values to which they were related in his value pattern. Note that these values are in a unique order for each person, so that wealth might be high for one, and low for another. Therefore the comments made on wealth would be recorded at various places in the patterns of different individuals, depending on where they had ranked the value "wealth." It follows that there would not necessarily be any scores recorded opposite the five highest values in a given student's pattern. For some students the five highest values seemed to attract most of the scores from their concept tests. For others there were actually no concept scores at the top of the value pattern, but many at lower points in the pattern. This reflects the fact that, while most students commented in terms of their highest values, some of them made comments which did not relate to their high values at all.

The final score was determined as follows. The plus and minus scores opposite each value were combined algebraically. Scores opposite the first value were weighted by five, those opposite the second value by four, those opposite the third value by three, those opposite the fourth value by two, and those opposite the fifth value by one. The five weighted scores were

then added algebraically. In order to keep away from negative numbers, 100 was added to each final score. Thus a score of 100 indicated that no comments had been made which concerned the first five values in the test, or that the plus and minus scores equalled each other and thus cancelled out. A score below 100 indicated a dominance of negative scores, and a score over 100 a dominance of positive scores.

There were then two variables, independently developed. The first was a score representing the student's attitude toward the proposition to abolish fraternities and sororities. The second was a score which represented the net balance between benefits and detriments believed by the student to be exerted upon him by a fraternity or sorority, and limited to those effects which were related to his highest values. One score is a reaction to a proposition, often described in the literature as a "degree of affect toward or against" the proposition. The other score represents conceptual content and its relationship to cherished values, or generalized conditions of living.

Data

The correlation between these two sets of variables was $-.804$, the negative sign being due to the fact that favorable attitude is shown by a low score, rather than a high one. In view of the generally accepted point of view that knowledge is not significantly related to attitudes (14), this is a striking result, although it is in conformity with the thesis that values are directive factors in behavior, and that attitudes are an aspect of behavior, and not causative factors.

The next question in the course of the study was whether a change in concepts would result in a change in attitude, and whether the change in attitude could be predicted from the change in concepts. During the three months following the administration of the tests, the students were supplied now and then with abstracted material from the literature on fraternities and sororities (1, 3, 4, 6, 7, 8, 9, 10, 12, 15, 16) and with an occasional opportunity to discuss them. Although most of the literature is favorable on a number of points, no position was taken by the class leader. In the meantime some of the students were having experiences of various kinds with

fraternities and sororities, and their concepts were being affected by these experiences.

At the end of the term the attitude tests and the concept tests were again administered. It was assumed on the basis of other studies (20) that the value patterns were not likely to have changed in such a short time. Seventy-nine students returned usable responses on this testing. The results of the final testing were handled in two ways. First a correlation was calculated, and turned out to be $-.73$, which is lower than the first, but still very substantial. Second, the changes in scores

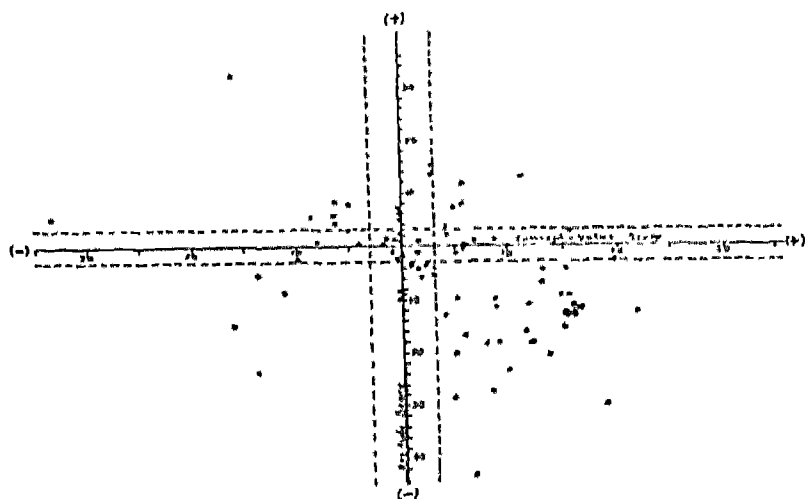


FIG. I. Relationship Between Changes in Scores from a Concept-Value Test and from an Attitude Test

were plotted graphically to reveal what relationships there were between such changes. Figure I contains the graph.

The first and fourth quadrants should contain all of the plotted points *if* an increase in the concept-value score always accompanied a change toward a more favorable attitude. Any reversals of this relationship would appear in the second and third quadrants. When all changes, no matter how small, were taken into account, 46 of the points are plotted in the first and fourth quadrants, and 18 in the second and third quadrants. These 64 students had changes in both scores. The other 15 students changed only one score, and were therefore plotted

on one of the lines between the quadrants. This provides strong evidence that changes in concepts tend to produce harmonious changes in attitudes. A question arises, however, as to what constitutes a significant change in scores. Since attitude scores were multiplied by ten to escape decimals, a change of three is really equal to a change of only .3 on the actual attitude scale, which varies for our students from 8.6 to 1.4, with a standard error of .175. A change of three points on the concept-value score may be judged in relation to the spread which goes from 61 to 153, with a standard error of 1.75. If a change of three points were considered to be insignificant, it would mean that 36 of the students failed to change one score by a significant amount. Of the remaining 43, 35 (or 81 per cent) were shown to be in quadrants one and four, and 8 (or 19 per cent) in quadrants two and three.

What may prove to be a significant lead in the relationship between concepts and attitudes is to be found in an examination of individual students' responses. There is a marked tendency for those with relatively neutral attitude scores to show no identifiable conceptual content about the subject. For example, contrast this quotation with the one quoted earlier as an answer to the same question:

Would not affect me in any way that I can possibly see. I am an independent so my life here is not connected at all with fraternities except in the indirect way that I know some men in houses and their abolition certainly wouldn't affect this in any way.

Again from the last question on the same student's sheet:

If they were abolished I think that the groups of men and women would still be living in these houses, even if taken over by the University, and would still form social groups—have functions together—as for example the girls dorms do now—open houses, etc. Of course probably these relationships would not nearly be so closely knit as now but otherwise I think life would go on much the same. Personally I can see no point in abolishing them. If a person don't like them he doesn't have to join and if he does like them he should not have the freedom of becoming part of one taken away from him.

This boy's attitude score was 64, which is neutral, and his concept-value score was 99, which means there was 1 weighted

negative point found on his concept check-list. It is suggestive of the fact that many individuals who show neutral attitudes toward objects or conditions do so because they have had no meaningful experiences with the object, and particularly *have acquired no meaning related to their most important values*. There were several responses similar to this example, but few responses in which *much* conceptual content was accompanied by a neutral attitude. In fact, the generalization seems warranted that extreme attitude scores occur in this study only when the person's highest values are involved in the proposition, *and* when there is a clear margin between negative and positive concept expressions.

In the case of two students, opportunity was accidentally afforded to attempt prediction of attitudes. Both of the students misinterpreted the instructions on the attitude test, and recorded their attitude toward fraternities rather than their attitude toward the abolition of fraternities. Suspicion that this had occurred arose when the two final sets of variables were being studied, because the scores for these two students were so greatly at variance with all the others. The experimenters, in discussing the two pairs of scores, decided to ask the students whether they had misinterpreted the instructions, and if so, whether they would complete the attitude test again in the proper manner. Both admitted they had misunderstood, and both took the test again. Prior to asking them about the matter, an attempt was made to estimate from their concept-value scores what the true attitude scores would be, by using the lists of paired variables of the other students as a rough equation table. In the case of one, the first attitude score was 3.3, and on the basis of his concept-value score it was estimated that his true attitude score should be between 8.0 and 8.3. On retest it turned out to be 7.9. In the other case, the first score was 8.4. The estimate indicated he should fall between 6.0 and 6.6. His retest score was 6.0. If such a study were to be repeated, it would be interesting to set up the relationships between attitude scores and concept-value scores for only half of the group at first; then from the concept-value scores of the remainder, attempt to predict the attitude scores before the attitude tests were scored. It seems probable that

in any fairly homogeneous group (speaking intellectually and socially) a scale could be developed from which a concept-value score could be used to estimate an attitude for other persons of similar background.

Conclusions

One's attitude toward a specific object or condition in a specific situation seems to be a function of the way one conceives that object from the standpoint of its effect on one's most cherished values. This means that while concepts alone cannot be shown to correlate highly with attitudes, when concepts are combined with subjective judgments as to the ability of the concept referent to help the individual achieve the things he wants, the basis exists for explaining attitudinal and emotional reactions. If, in the judgment of the individual, a given object has no effect on his high values, he will exhibit a neutral attitude toward it. If he conceives it to be destructive of his high values he will exhibit a negative attitude toward it, and vice versa.

Attitudes will change when either one of the two basic factors change. If the value pattern is altered, it follows that certain objects which might have promoted formerly cherished values, but which do not promote presently cherished values, will no longer evoke the pleasant attitudes they formerly evoked. Also, if the concept of a given object or condition is changed, so that its potential effect on certain values is believed to be altered, the attitude will reflect that conceptual change. Unless the nature of the change in values or in concepts is known, the change in attitude will be relatively unexplainable.

Since value patterns appear to be fairly resistant to change, it seems likely that most changes produced in attitudes will be brought about by making changes in the concept of the object toward which the attitude is expressed. This emphasizes the role of education in producing changes in behavior, but offers a realistic explanation for the fact that educational practices which do not really change concepts, and which do not take into account value patterns, are futile.

One of the most needed psychometric devices, and one which is almost nonexistent at the present time, is a test or a set of

tests or techniques for measuring functional concepts. It is suspected that these devices will not need to be as subtle and indirect as has been generally supposed, and that the futility of direct tests of ideas has been caused by failure to take into account all of the directive factors in behavior. An individual can probably give a fairly valid verbal description of his motives, especially his concepts of means and processes. When his values are known, these verbally expressed means and process concepts are likely to prove most enlightening as to the reasons for his behavior.

REFERENCES

1. Bogardus, E. S. "The Fraternity As a Primary Group." *Sociology and Social Research*, XXIV (1940) 456-460.
2. Cottrell, Jr., Leonard S. "The Analysis of Situational Fields in Social Psychology." *American Sociological Review*, VII (1942) 370-387.
3. Crocker, Lionel. "Fraternities and Democracy." *Vital Speeches*, IX (1943) 766-768.
4. Cronbach, L. J. "Stereotypes and College Sororities." *Journal of Higher Education*, XV (1944) 214-216.
5. Cutler, Virginia F. "The Development of An Educational Instrument To Reveal Personal and Family Values in the Choice of a Home." Ph.D. Thesis, Cornell University, 1946.
6. Deakins, C. E. "In Defense of Fraternities." *Journal of Higher Education*, XII (1941) 259-264.
7. Lehman, H. C. "College Marks and the Fraternity Pledge." *Journal of Applied Psychology*, XIX (1935) 9-28.
8. McCreery, O. C. and Mott, G. F. "Seeing Fraternities in a Larger Frame." *Journal of Higher Education*, IX (1938) 331-334.
9. Porterfield, A. W. "Hereafter of College Fraternities." *School and Society*, LXI (1945) 193-196.
10. "Relations Between College and Fraternity." Association of American Colleges and National Interfraternity Conference. *School and Society*, XLVII (1938) 253-255.
11. Richardson, H. M. "Community of Values as a Factor in Friendships of College and Adult Women." *Journal of Social Psychology*, XI (1940) 303-312.
12. Rigg, Melvin. "Motivation in the Arts College." *Journal of Higher Education*, VIII (1937) 101-102.
13. Scheaffer, B. R. "The Validity and Utility of the Allport-Vernon Study of Values Test." *Journal of Abnormal and Social Psychology*, XXX (1936) 419-422.
14. Smith, George H. "A Note on Attitudes and Information." *Journal of General Psychology*, XXXVII (1947) 193-198.

15. "Survey of Social Life at Dartmouth." *School and Society*, XLIII (1936) 664.
16. Turner, Fred and Tunis, J. R. "Are Fraternities Worth While?" *Rotarian*, LI (1937) 2-4, 14-17.
17. Woodruff, Asahel D. *A Study of Choices*. A test of personal values. 311 Stone Hall, Cornell University, Ithaca, N. Y. 1948
18. ——. "A Study of the Directive Factors in Individual Behavior." Ph.D. Thesis, The University of Chicago, 1941.
19. ——. "Personal Values and the Direction of Behavior." *School Review*, I. (1942) 32-42.
20. ——. "The Relationship Between Functional and Verbalized Motives." *Journal of Educational Psychology*, XXXV (1944) 101-107.

MEASUREMENT IN RURAL HOUSING—A PROGRESS REPORT

CHARLES L. DURRANCE, JR.
University of Florida

WHAT can schools do to improve housing in a community? How can schools help people to use the scientific knowledge that is available? Is it the responsibility of the school to provide a functional program that will include activities in all grades to help pupils repair and improve their homes, beautify the grounds, and eliminate health hazards?

The Florida Project

The Project in Applied Economics at the University of Florida is vitally interested in these questions.

From grants-in-aid from the Alfred P. Sloan Foundation, the Project, established in 1940, is a part of a nationwide program to improve the level of living. It operates under the direction of a committee composed of representatives from the College of Education, the Administration of the University, and the State Department of Education. It has as its purpose the development of housing materials and activities for all grades.

The Project attempts to find out whether or not the houses of all boys and girls will be improved as a result of the instructional program in housing. Constant checks are made to see if the classroom activities are extending into the homes.

The Florida Project works with three assisting schools, selected because they are located in typical rural, nontourist communities and because the faculties were interested in experimenting with new curriculum materials to improve living in the communities. These schools have agreed to center attention on experiences designed to improve housing. Paired with these assisting schools are three control schools, which have agreed to cooperate in a measurement program but not to use the Project materials.

Getting Started

Before the Project could make much headway, it was necessary to develop a housing inventory to determine what were the greatest housing needs of the assisting and control communities and which needs the schools should use for developing a program of housing education. This inventory included 96 items, which may be classified under the following general groupings: (1) size of house and arrangement; (2) roof, frame, foundation, steps; (3) interior—walls, floors, ceiling; (4) sleeping space; (5) kitchen—space and facilities; (6) ownership—valuation and attitudes; (7) yard and building site; (8) heating facilities and fire hazards; (9) lighting—natural and artificial; (10) exterior walls; (11) doors and windows; (12) water supply and sanitation; (13) age of house.

A staff of workers used this inventory to survey housing conditions. Every house in the assisting and control communities was photographed and conditions of the houses were indicated. This information is now serving as a basis for the evaluation of the housing instruction program to see whether conditions have been improved.

Results of the Seven-Year Study

Significant results are reported on only two classes of homes: (1) The same house, same occupant both years and (2) the same house surveyed in 1940 and 1947, even though the occupant changed. Some of the more significant findings are here presented.

An examination of the summary of results indicates that in 24 selected items the experimental communities made the greatest improvement in housing conditions. In six items the control communities made the greatest improvement. In ten items there was no significant difference between the two groups.

The experimental communities made more improvements than the control communities in the elimination of fire and health hazards, especially in improving the water supply, eliminating the unsanitary surface privy, wiring for electricity, substituting adequate heating facilities, using some means of

refrigeration, making repairs on the roof, cleaning of yards littered with garbage, screening windows and doors, and building adequate chimneys and flues.

The control communities made the greatest improvement in termite proofing and in the addition of village garbage disposals. It will also be noted that there were fewer homes in the control communities with screens and steps in worse condition in 1947 than there were in 1940.

There was no significant difference between the experimental and control communities in the number of new houses built since 1940. A very high per cent of the houses in both groups still have no other means of heating than the open fireplace.

The influence of the school program on the changes in the community has not been determined. Changes in the community were not as great as might be expected, partly because the total school program was not involved. Not one of the three experimental schools had the same principal for all of the seven-year period. There were five different principals in one school, four in another, and three in the third school. The turnover of teachers in the three schools followed the same pattern. In one school the average yearly turnover of elementary teachers was 21 per cent; in the second school it was 75 per cent; and in the third school it was 46 per cent. In 1946-1947 all of the elementary teachers in one school were new. In one of the three high schools, where 38 teachers were employed during the seven-year period, 26 of the 38 teachers taught only one year.

Findings of Preliminary Housing Survey in 1940

In defining the term "community" a different interpretation was used from the one generally accepted. "For the purposes of this study, a community is defined as consisting of all those families, and only those families, which send at least one child to a specific school being studied."¹

There were 715 houses surveyed in the six rural Florida communities, with these significant findings:

¹ Mosier, Charles I. "Measurement in Rural Housing—A Preliminary Report," *EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT*, II(1942), 139-152.

Water supply—61 per cent had only outside hand pumps or wells
 Toilet facilities—16 per cent had no toilet facilities; 70 per cent had no better than an open surface privy
 Lighting—72 per cent had no electricity
 Termite damage—24 per cent showed evidence of damage
 Termite proofing—only 2 per cent of homes were termite-proofed
 Sink—80 per cent had no kitchen sink whatever
 Refrigeration—52 per cent had no refrigeration
 Garbage disposal—15 per cent had yards littered with garbage and refuse
 Screens—41 per cent had all outside openings screened
 Cabinet space—15 per cent had no pantry or storage space in kitchen
 Chimney and flues—63 per cent constituted a fire hazard
 Shrubby—37 per cent had no shrubs around the house
 Roof conditions—47 per cent needed some repair

Dr. Mosier indicates that the differences between the experimental and control groups were small, but that the control group showed a slight superiority.

Significant Changes in Houses from 1940-1947

From October of 1946 to November of 1947 a staff of workers gathered data for the second phase of the evaluation program of the University of Florida Project in Applied Economics.

All homes in the three experimental communities and the three control communities (where there were children in school) were visited and conditions of the houses were indicated on the *Questionnaire for the Partial Measurement of the Effectiveness of a School Program in Housing Instruction 1946-47*.

There were 823 homes surveyed in 1947—an increase of 108 over 1940. These 823 homes were then divided into seven classifications:

1. Same house, same occupant both years
2. Same house, same occupant, but no children in school in 1947
3. Same family, different house
4. Same house, different occupant
5. Old house—not surveyed in 1940 because no children in school, but family does have children in school in 1947
6. New house (built since 1940 and no 1940 house available for comparison)
7. House surveyed in 1940 but not in 1947 (with reason given)

For the tabulation of data, two tables were made to indicate changes made during the seven years. These two tables were: Table 1—same house, same occupant both years (classifications 1 and 2 above); Table 2—changes in same houses surveyed both years (classifications 1, 2, and 4 above). It was believed that information about classifications 3, 5, 6, or 7 would not give significant data.

The comparison of the number of houses tabulated for these two tables follows:

Town	Same House, Same Occupant	Same Houses Sur- veyed 1940 and 1947
Bardin.	14	25
Ladyland.	18	28
Summerfield.	53	75
Melrose.	41	66
Chiefland.	114	144
Bransford.	111	161
Total Experimental Group	181	244
Total Control Group	170	245

For purposes of simplification, the experimental communities will hereafter be referred to as Group "E"; the control communities as Group "C."

Some of the significant findings are presented here (information obtained from Tables 1 and 2):

Water Supply.—Table 1 indicates that in the experimental communities (1940) 120 homes had an outside hand pump and 84 of the homes in the control group depended on the outside hand pump for water. By 1947 this decreased to 84 for the experimental group and to 64 for the control group. Thus in 1947, 46.4 per cent of the homes in the experimental group and 37.7 per cent of the homes in the control group still had nothing better than a hand pump. The percentage decrease, however, favored the experimental group 20 per cent to 12 per cent. In the experimental group the number of power pumps increased 206 per cent during the seven-year period and 100 per cent for the control group. Table 2 indicates that fourteen homes in both groups changed from depending on a well or carrying water to the use of a power pump and tank.

Toilet Facilities.—Although the experimental communities

TABLE 1
Comparison of Selected Items in Same Houses and Same Occupants

	Total "E" '40	'47	Total "C" '40	'47
Rebuilt.				
Yes	55	34	55	69
No	124	147	106	101
Water Supply:				
Well	7	2	20	5
Outside hand pump	120	84	84	64
Power pump and tank	32	94	48	96
Toilet Facilities:				
None	20	8	27	14
Surface privy	100	77	92	81
Sanitary pit privy	31	17	8	6
Flush toilet	25	42	25	55
Lighting:				
Old-fashioned lamps	97	45	114	61
Electricity	66	133	42	100
Heating Facilities:				
None except cookstove	9	2	2	2
Fireplace	113	91	125	112
Heating stove	11	67	19	19
Heating stove and fireplace	11	22	12	32
Kind of cookstove:				
Small wood stove	75	28	69	38
Large wood range	33	41	62	38
Electric	0	9	1	13
Gas	2	22	5	27
Kerosene	35	94	23	53
Is There a Sink?				
None	153	123	119	94
Yes; no outside drain	5	3	6	21
Outside drain and water supply	22	54	36	55
Termites:				
Extensive damage	11	2	10	4
Obvious evidence	23	33	22	37
No protection, no evidence	142	137	117	111
Termite proofing	5	9	1	18
Kind of Refrigeration:				
None	87	43	83	52
Ice box, no ice	7	4	7	9
Ice box, with ice	40	58	35	33
Mechanical refrigerator	45	76	36	74
Garbage Disposal:				
Yard littered	42	7	26	19
Yard clean	135	162	135	126
Covered cans	4	11	0	2
Village disposal	0	0	0	13
Screens:				
No screens	41	41	57	50
Some screens	51	25	32	30
All good screens	88	116	72	91

TABLE 1—*Concluded*

	Total '40	Total "E" '47	Total '40	Total "C" '47
Water Supply:				
Very poor	5	0	1	0
Poor	89	48	37	22
Fair	11	3	4	16
Good	65	91	112	138
Cabinet Space:				
Completely absent	4	0	3	9
Poor makeshifts	16	10	11	18
Inadequate	68	123	44	54
Adequate-excellent	92	48	107	85
Chimney and Flues:				
Stove pipe	85	26	46	44
Poorly insulated	14	14	20	15
Well insulated	80	137	90	86

had eight more homes with just surface privies in 1940, they had four less than the control communities in 1947. Adequate toilet facilities (either sanitary, pit privy or flush toilet) were in 32.6 per cent of the experimental homes and 35.9 per cent of the control homes in 1947. However, 17 "E" and 13 "C" changed from no toilet in 1940 to a surface toilet in 1947; and 30 "E" and 13 "C" changed from a surface privy to a sanitary pit privy. Unfortunately, four homes in the "E" group which had a surface privy in 1940 had no toilet in 1947.

There was a general decrease in the number of homes with no toilet facilities. Group "E" had a decrease of 12, with 8 homes still without toilets. Group "C" had a decrease of 13, with 14 homes with no toilet facilities.

Lighting.—Sixty-seven homes in Group "E" have installed electricity; 58 in Group "C." If we look at the comparison of the same houses surveyed both years (Table 2), 89 (36.5 per cent) of the homes in Group "E" have added electricity; 73 (30 per cent) in Group "C." Sixty-one homes in Group "C" still have old-fashioned lamps; 45 homes in Group "E." Two homes in Group "C" have even changed from electricity to old-fashioned lamps.

Termite Damage.—Damage from termites remained the same in both years in the homes of Group "E" and increased from 20 to 24 per cent in Group "C." With reference to termite proofing, the increase was greater in the control group—17 as compared to four.

TABLE 2

Significant Changes in Houses (Comparison of Same Houses Surveyed in 1940 and 1947)

	Total "E"	Total "C"
Rebuilt or Remodeled.		
Yes	50	99
No	194	146
Water Supply:		
Well to porch pump	0	2
Hand pump or carried to power pump	74	71
From well or carried to hand pump	9	5
Toilet Facilities.		
None to flush toilet	2	3
No toilet to surface privy	17	13
No toilet to pit privy	7	3
Surface privy to pit privy	30	13
Surface privy to inside toilet	12	18
Pit privy to inside toilet	8	9
Surface privy to no toilet	4	0
Pit privy to surface privy	8	5
Lighting:		
Electricity added	89	73
Old-fashioned lamps to improved	6	4
Electricity to old-fashioned lamps	0	2
Old-fashioned lamps to candles	0	1
Termites:		
No protection to obvious evidence and damage	30	32
Damage from termites to no evidence	34	15
Termite proofing added	7	20
Chimneys and Flues:		
Stove pipe to brick insulated	71	8
Brick insulated to poorly insulated	3	5
Brick insulated to stove pipe	6	12
Brick poorly insulated to insulated	16	9
Stove pipe to poorly insulated	8	5
Stove pipe to none	3	3
Some insulation to none	5	9
Kind of Cookstove:		
Wood stove or range to kerosene	106	44
Electric stove added	10	12
Gas stove added	20	27
Small wood stove to large wood range	10	6
Kerosene to wood stove	2	1
Electricity to kerosene	0	1
Is There a Sink?		
No sink to outside drain and water supply	29	23
No sink to water supply, no outside drain	2	20
Outside drain to no sink	2	4
No outside drain to outside drain	5	4
Refrigeration:		
No refrigerator to ice box	53	21
No refrigerator to mechanical refrigerator	17	23
Ice box to mechanical refrigerator	22	20
Mechanical refrigerator to ice box	2	1
Ice box or mechanical refrigerator to none	17	9
Ice box no ice, to ice box with ice	2	3

TABLE 2—*Continued*

	Total "E"	Total "C"
Garbage Disposal:		
Yard littered to yard clean	55	17
Yard clean to littered	13	9
Covered garbage cans added	7	4
Village disposal added	0	22
Screens:		
No screens to all good screens	16	7
No screens to some good screens	9	6
Poor screens to some good screens	26	8
Some screens to poor or no screens	27	18
(worse than in 1940)		
Some good screens to all good screens	9	3
Steps:		
From good to fair	21	31
From poor to fair	13	10
From good to poor	11	1
From fair to good	34	14
From fair to poor	15	4
Steps added—from no steps to poor	7	0
Steps added—from poor to good	21	5
Yard:		
From clean to littered	14	12
Lawn added	35	19
Littered to clean	13	2
Swept or raked to clean, no grass	1	9
From lawn to no lawn	15	9
From grass cut to littered	4	1
From clean to raked, but no grass	5	0
Shrubs:		
Inadequate to none	35	21
Adequate to inadequate or none	9	7
Shrubs added	31	26
Well-placed, nor enough to inadequate	11	2
Inadequate poorly placed to well-placed	13	3
Finish of Floors:		
Linoleum added	82	46
Added living room rug	33	6
Conditions worse than in 1940	10	19
Other improvements made	9	1
Roof Conditions:		
Repairs made on roof	85	38
Roof deteriorated since 1940	47	39
Kitchen Arrangements:		
Improvements made	29	13
1947 conditions worse than in 1940	123	78
Cabinets:		
Improvements made	47	22
1947 conditions worse than in 1940	98	56
Condition of Water Supply:		
From good to fair	0	17
From good to poor	23	6
From poor to fair	2	0
From poor to good	42	20
From fair to poor	7	0
From fair to good	10	2

Heating Facilities.—Both groups have added heating stoves or circulating heaters (an increase of 67 in Group "E" and 10 in Group "C"). However, there are still a large number of homes depending entirely on fireplaces for heating facilities (91 in Group "E" and 112 in Group "C"). Thus, 50 per cent of the homes in the experimental communities and 66 per cent of the homes in the control communities still use only fireplaces for heating. Group "E" made more improvement in this than Group "C" (a decrease of 12 per cent in Group "E" and 8 per cent in Group "C"). There are still four homes in the six communities which have no heating facilities except a cookstove.

Kind of Cookstove.—Table 1 reveals that there has been a great shift from the small wood stove and wood range to the use of kerosene for cooking. In 1940, 108 homes (60 per cent) in Group "E" and 131 homes (77 per cent) in Group "C" depended on wood. By 1947 this had decreased to 69 homes (38 per cent) in Group "E" and 76 homes (45 per cent) in Group "C." Between 1940 and 1947 Group "E" has added nine electric stoves and Group "C" 12. Both groups are practically equal in the addition of gas (butane) stoves: Group "E" increased from two to 22 and Group "C" from five to 27.

Table 2 indicates the same change from wood stoves to kerosene when we consider the same houses surveyed both years. In the experimental communities 106 of the 244 families changed from wood to kerosene for fuel, and 44 of the 245 families in the control communities made this change. In the addition of gas and electric stoves the increase in both groups parallels that revealed in Table 1. People who used kerosene in 1940 continued to use it in 1947, for only three switched from kerosene to wood (two in Group "E" and one in Group "C").

Is There a Sink?—There is still a high percentage of homes without a kitchen sink. Of the homes surveyed in 1940, 84.5 per cent of those in Group "E" and 70 per cent of the homes in Group "C" had no sink. By 1947, 68 per cent of Group "E" and 55 per cent of Group "C" still had no sink. Thus both groups made almost identical improvement (16.5 per cent for Group "E" and 15 per cent for Group "C"). The experimental group, however, made more improvement in the addition

of a sink with outside drain and water supply: 32 homes in Group "E" and 19 homes in Group "C."

Table 2 reveals these significant changes: (1) 29 homes in Group "E" and 23 homes in Group "C" which had no sink in 1940 had added a sink and water supply by 1947, (2) four homes in Group "C" and two homes in Group "E" had a sink in 1940 but none in 1947.

Refrigeration.—Table 1 shows that in 1940 Group "E" had 87 or 48 per cent of the homes without refrigeration; Group "C", 83 homes, or 49 per cent. By 1947 this number had decreased to 43 or 23.8 per cent in Group "E" and to 52 or 31 per cent in Group "C." Thus Group "E" made 24 per cent improvement, while Group "C" made only 18 per cent improvement. Group "E" added 31 mechanical refrigerators and Group "C" added 38.

Table 2 indicates the substitution of mechanical refrigerators for ice boxes, probably the result of the introduction of Rural Electrification Administration lines into the communities. In many homes the survey revealed electrical conveniences, such as refrigerators, stoves, and washing machines, where there was no evidence of preservation of the house in such matters as painting, roofing, or protection against termites. If we look at the addition of some kind of refrigeration, 26 more homes in Group "E" than in Group "C" made this change. This is evidenced by the fact that Group "E" added 53 ice boxes, while Group "C" added only 21. Unfortunately, 17 homes in Group "E" and 9 homes in Group "C" having refrigeration in 1940, had none in 1947.

There were 45 homes in Group "E" and 36 homes in Group "C" with mechanical refrigeration in 1940. Of this group of 81 only three families changed in 1947 to the use of an ice box.

Garbage Disposal.—There was a great decrease in the number of homes with yards littered with rubbish and garbage. The greatest improvement was made in the experimental communities. The decrease in littered yards here was 35 (from 42 to 7), while the decrease was 7 (from 26 to 19) in the control communities. Most of the families either burned or buried or fed garbage to the animals. Thus there were very few covered garbage cans. One significant change to be noted during the

7-year period is that the town of Branford (Group "C") made available village garbage disposal and that 13 families (same house in 1940) used this method of garbage disposal.

Table 2 shows that Group "E" also made the greatest improvement in cleaning the yards—55 as compared to 17 for Group "C."

Screens.—Very little improvement seems to have been made in decreasing the number of homes with no screens. The number of homes without screens remained the same in Group "E" (41), while it decreased seven for Group "C" (57 to 50). However, if we look at the number of homes with all good screens the number increased 28 in Group "E" and 19 in Group "C." Looking at Table 2, we see that 16 homes in Group "E" without screens in 1940 added all good screens. Only seven homes in Group "C" made this improvement. On the other hand, if we compare the condition of screens in the homes for both years, 27 homes in Group "E" and 18 homes in Group "C" had screens that were in worse condition in 1947 than in 1940.

Cabinet Space.—Here the experimental communities had more improvement than the control communities. In 1940, four homes in Group "E" and three homes in Group "C" had no cabinet space whatsoever. In 1947 all homes in Group "E" had some cabinet space, while the number without cabinets actually increased in Group "C" from three to nine. The same is true of the homes with "poor makeshift" cabinets. In Group "E" the number decreased from 16 to 10, while in Group "C" it increased from 11 to 18.

Both groups seemed to do little about providing adequate cabinet space. In fact, there were more homes with inadequate cabinet space in 1947 than in 1940. Table 2 reveals that while in Group "E" the cabinet space conditions were worse in 1947 than in Group "C," more improvements were made than in Group "C" (47 as compared with 22).

Chimney and Flues.—In 1940, 85 (47 per cent) of the homes in Group "E" and 46 (27 per cent) of the homes in Group "C" had flues that were definitely a fire hazard. The greatest improvement in this respect during the 7-year period was made in Group "E"—a decrease to only 26 (14 per cent) in 1947 as compared to 44 for Group "C" (21 per cent). Group "E" had

fewer well-insulated chimneys in 1940, but in 1947 it had a greater number of well-insulated chimneys than did Group "C"—an increase of 57 well-insulated chimneys for Group "E" and a decrease of four for the control group.

Table 2 gives a clearer picture of this trend

Significant Change (1940-1947)	Group "E"	Group "C"
From stove pipe to brick insulated	71	8
From poorly insulated to well insulated	16	9
From stove pipe to poorly insulated	8	5

Thus we see that 95 homes in Group "E" made an effort to eliminate fire hazards, while only 22 homes in the control group accomplished this result. It will be noted also in Table 2 that 12 homes in Group "C" with brick insulation in 1940 reverted to a stove pipe in 1940; this happened to only six homes in the experimental group.

Steps.—Improvements were made in 75 homes in Group "E" and only 29 homes in Group "C." On the other hand, in 1947 conditions were worse in 47 of the Group "E" homes and 36 of the Group "C" homes than they were in 1940.

Miscellaneous Changes.—Lawns were added to 35 homes in Group "E" and only 19 in Group "C."

Addition of shrubbery was almost equal in both groups—31 for Group "E" and 26 for Group "C."

Two changes stand out in the finish of the floors: (1) addition of linoleum, (2) addition of living room rug. In each of these the experimental group had the advantage (see chart below).

Significant Changes (1940-1947)	Group "E"	Group "C"
Linoleum added	82	46
Living room rug added	33	6
Conditions worse than in 1940	10	19
Other improvements made	9	1

It will also be noted that conditions in 1947 were worse in almost twice as many homes in Group "C" as in Group "E."

Almost the same evidence holds true for repairs on the roof. Eighty-five homes in Group "E" and 38 in Group "C" made repairs on the roof in the period between 1940 and 1947. However, in 39 homes of Group "C" and 47 homes of Group "E" the roof conditions deteriorated during the 7-year period.

So far as making improvements in the kitchen arrangement

is concerned, not much can be said. Only 29 in Group "E" and 13 in Group "C" made any improvement along this line.

The tables indicate that much progress has been made in the improvement of housing conditions in the six communities surveyed in 1940 and 1947. The table below helps us to see (1) selected items in which the experimental group excelled; (2) selected items in which the control group excelled, (3) selected items in which there was no appreciable difference.

Items in which experimental group made greatest improvement

	Group "E"	Group "C"
Decrease in per cent of homes depending on outside hand pump	20	12
Increase in per cent of homes adding power pump and tank	206	100
Increase in number of homes changing from surface privy to sanitary pit privy	30	13
Increase in per cent of homes adding electricity	37	30
Number of homes still depending on old-fashioned lamps	45	61
Increase in per cent of homes damaged from termites006	.04
Addition of circulating heaters or heating stoves	67	10
Increase in number of homes substituting kerosene for wood stove for cooking . .	106	44
Number of homes adding kitchen sink with outside drain and water supply . .	32	19
Decrease in per cent of homes without refrigeration	24	18
Number of homes adding some kind of refrigeration	70	44
Number of homes making improvements in kitchen arrangement	29	13
Number of homes making repairs on roof	85	38
Decrease in number of homes with yards littered with garbage	35	7
Increase in number of homes with all good screens	28	19
Number of homes with all good screens in 1947 but with no screens in 1940 . .	16	7
Decrease in number of homes with no kitchen cabinet space	4	(increased 3 to 9)

Number of homes making some improvement in cabinet space.	47	22
Number of homes eliminating fire hazards in flues	95	22
Number of homes changing from stove pipe to well-insulated chimney	71	8
Number of homes making improvements in steps.	75	29
Number of homes adding lawns.	35	19
Number of homes adding linoleum to unfinished floors	82	46
Number of homes adding living room rug	33	6

Items in which control group made greatest improvement

Number of homes adding termite proofing.	17	4
Decrease in per cent of homes depending on wood stove for cooking.	22	32
Number of homes with refrigeration in 1940 but with no refrigeration in 1947	17	9
Number of homes adding village disposal for garbage	0	13
Number of homes with screens in worse condition in 1947 than in 1940.	27	18
Number of homes with steps in worse condition in 1947 than in 1940.	47	36

Items in which there was no significant difference between the two groups

Number of families building new houses since 1940.	69	58
Decrease in per cent of homes with no toilet facilities.	7	8
Per cent of homes with adequate toilet facilities.	33	36
Number of homes still depending on fireplace as only means of heating house	91	112
Number of homes with no heating facilities except cookstove.	2	2
Number of homes adding electric stoves.	9	12
Number of homes adding gas stoves	20	22
Number of homes substituting wood stoves for kerosene in cooking.	2	1
Decrease in per cent of homes with no kitchen sink in 1947.	17	15
Number of homes adding mechanical refrigerators.	31	38

Implications Arising from this Study

To what extent the school influenced these changes recorded in this survey information is not available. It must be as-

sumed, however, that in some instances where the changes were much greater in the experimental communities than in the control communities, education must have had some influence on these changes.

It must be pointed out also that some changes were not as significant in the experimental groups as might be expected because of the rapid turnover of principals and teachers. A real program in housing instruction was not adhered to, and the information was imparted not as a total school program but in most cases in individual classrooms. Teachers changed so frequently that by the time a teacher became familiar with the program she moved to another area. This necessitated beginning again working to interest the new teacher who had taken her place. It is believed that if the experimental schools had undertaken the project wholeheartedly, in a total school program directed by an energetic, community-minded principal, the significant changes would be much more in evidence.

Many of the changes came as a result of the introduction of Rural Electrification Administration into the communities. Other improvements came from high wages during the war years and the desire to have the comforts of life.

This all points up the necessity for all community agencies to work together for raising the economic level of living in the community. It indicates the need also for teacher-training institutions to prepare teachers to select their instructional material and to provide activities so that boys and girls will be able to understand and to meet their basic economic needs. When teachers with this kind of training take their place in the schools, more schools will have as an integral part of their program activities for the improvement of living. Communities will then secure more adequate housing through education.

INTELLIGENCE TEST RESULTS OBTAINED FROM A SPECIFIC TYPE OF ARMY A.W.O.L.

JERRY H. CLARK

University of California, Santa Barbara College

At a Center which processed soldiers who had gone A.W.O.L. from a unit prior to its shipment overseas, 100 randomly selected literate trainees were administered four verbal sub-tests of the *Wechsler Mental Ability Scale, Form B*. It is pointed out that these trainees had committed a serious offense in terms of the effect on the war effort—they had deserted their organizations at the time when they were most needed. For this reason, this offense is a specific type of A.W.O.L. and can be viewed as a very serious infraction of Army discipline. The following information concerning the 100 trainees comprising the experimental group in this study is submitted: mean age, 24.72 years; mean years of school completed, 9.19; mean weekly wage received prior to induction, \$41.00; and mean number of weeks of Army service, 115.

The *Mental Ability Scale, Form B*, was developed by Wechsler for clinical use in the Army, and was one of the authorized Army tests. Four of the verbal sub-tests (Information, Arithmetic, Comprehension, and Similarities) were orally administered in order to obtain a measure of the mental ability of these trainees. The distribution of standard scores based on these four sub-tests is presented in Table 1. The range of this distribution is 55-150, the mean standard score is 87.10, standard deviation, 18.20. A standard score of one hundred would be the strictly "normal" score. Seventy-six per cent obtained scores below this level.

These intelligence test scores may also be interpreted in terms of the IQ. For this group of 100 trainees, Table 2 presents the distribution of IQ's. The range of this distribution is 71-135, the mean is 92.65, the standard deviation is 11.95. These IQ scores, presenting in another fashion the scores obtained on

the four verbal sub-tests, also reveal a skewness in the lower end of the distribution. Only 29 of the trainees have IQ's of 100 or above. The division of test scores according to Wechsler's categories (2) was tabulated as shown in Table 3.

This table reveals that there are gross differences in the

TABLE 1
Standard Scores: Wechsler Mental Ability Test, Form B for 100 A.W.O.I.'s

Standard Score	Frequency
140 -	1
130-139	0
120-129	4
110-119	6
100-109	13
90-99	20
80-89	16
70-79	22
60-69	16
50-59	2
	100

TABLE 2
Intelligence Quotients: Wechsler Mental Ability Scale, Form B

IQ	Frequency
130 -	1
125-129	0
120-124	0
115-119	4
110-114	6
105-109	3
100-104	15
95-99	15
90-94	12
85-89	14
80-84	16
75-79	12
70-74	2
	100

classification levels between the percentages expected according to Wechsler (in a normal distribution) and on the percentages obtained from a tabulation of the one hundred scores obtained in the Processing Center. Of the 100 cases, 25 should have scored above the "normal" bracket; in actuality, only eight are so classified. Again, the expectancy below the "normal" bracket would have been 25; actually, 48 scored in this range. The number of cases in the "dull normal" and "border-

line defective" categories is thus about twice the normal expectancy. Although it is noted that none of these 100 trainees would be classified in the "defective" bracket, this may probably be accounted for in terms of the selection of the group which specified fourth-grade literacy in order that the MMPI

TABLE 3
Distribution According to Intelligence Classification

Classification	IQ Limits	Processing Center Frequencies	Per Cent Included	Expected Percentages According to Wechsler
Very Superior	128 and over	1	1	2.2
Superior	120-127	0	0	6.7
Bright Normal	111-119	7	7	16.1
Normal	91-110	44	44	50.0
Dull Normal	80-90	34	34	16.1
Borderline	66-79	14	14	6.7
Defective	65 and below	0	0	2.2
		100	100	100.0

TABLE 4
Distribution of Sub-test Scores: Wechsler Mental Ability Scale, Form B

Sub-test	Frequency	Percentage	Percentage	Percentage
15	1	3	1	
14	1	0	1	
13	2	7	0	5
12	8	9	8	8
11	10	3	6	5
10	10	15	13	12
9	18	19	16	20
8	24	18	15	21
7	17	2	13	15
6	3	15	11	14
5	6	6	12	
4		2	2	
3		1	2	
Total	100	100	100	100
Mode	8	9	9	8
Mean Score	8.82	8.87	8.16	8.73
S.D.	2.07	2.59	2.41	1.96

might be administered. From an inspection of the mean score, the standard deviation, and the classification levels in the above table, it is seen that this group is significantly lower in intelligence than the standardizing population used by Wechsler. The distribution of the scores of the 100 trainees on these four sub-tests is presented in Table 4.

It is seen that the mean score obtained on the Information sub-test is 8.82; twenty-two of the group scored above 10 and 68 below 10. On the Arithmetic sub-test the mean is 8.87, with 22 scoring above 10 and 63 below 10. The mean of the Comprehension sub-test scores is 8.16, sixteen obtained a standard score above 10 and 71 below 10. On the Similarities the mean is 8.73; eighteen scored above 10 and 70 below 10. By inspection, it appears that the distribution of the Comprehension sub-test scores is the only one which might be significantly different from the distributions of the other sub-tests. For this reason, the differences are presented in Table 5 in terms of the critical ratio.

It would seem, therefore, that the Comprehension scores are

TABLE 5
Comparison of Means: Comprehension with Information, Arithmetic, Similarities

Scales	Means	D.	S.D. of D.	Product- Moment r	C.R.
Information-Compre- hension	8.82-8.16	+ .66	.182	+ .68	3.63
Arithmetic-Compre- hension	8.87-8.16	+ .71	.279	+ .38	2.54
Similarities-Compre- hension	8.73-8.16	+ .57	.254	+ .38	2.24

significantly lower than those of Information, Arithmetic and Similarities. The critical ratio of 3.63 of the difference between Information and Comprehension fulfills the demands for reliability; the critical ratios with respect to Arithmetic and Similarities are well above the 5 per cent level of confidence. Pattern analyses of the Wechsler are generally not based on the administration of only four sub-tests; however, this does not mean that these significant differences between sub-tests found above are meaningless or without interpretation. In discussing the Comprehension sub-test, Wechsler states that it might be termed a test of common sense and that success on this test apparently is dependent upon a certain degree of practical information and a "general ability to evaluate past experience"(2). This "common sense" interpretation and "evaluation of past experience" might well apply to the group under discussion. It is not stretching the facts to say that

"good common sense" would provide a soldier with the knowledge that such behavior as "boat-jumping" (going A W O L to avoid overseas service) would really not solve his problem and that eventually he would have to suffer for the error of his ways. Moreover, the fact that 55 of these soldiers had been A.W.O.L. before might indicate an inability to evaluate past experience properly (Admittedly, this fact might have other interpretations) Twenty-eight admitted being A W.O.L. three times or more, including the present offense.

This lower score, obtained on the Comprehension test, might also be better understood if the items of the sub-test are examined. Inspection reveals that many of the items might be based on the degree of "social intelligence," as they have to do with fair play and one's responsibility to his fellow man and to society. For example, consider the question, "What should you do if someone smaller than you started to pick a fight with you?" This question was number two on the list, and apparently should not be difficult except for the very dull, however, it was surprising to see how many "o" or "1" responses were obtained in comparison to the number of "2" (highest score possible) responses. Many of these soldiers responded with "beat him up," or "let him have it." Many of the other questions having to do with social understanding also seemed to present great difficulty. Although it might be said that these prisoners violated Army rules and frequently repeated the offense just because the error of their ways was never explained to them, there seems to be a possibility that they were actually deficient in social intelligence and that they were actually incapable of properly evaluating past experience. This brings to mind one of the chief characteristics of the psychopath as delineated by Cleckley (1)—his inability to profit by experience, regardless of how chastening his experiences may be.

Summary

1. Four verbal sub-tests of the *Wechsler Mental Ability Scale, Form B*, were administered to 100 A W O L. soldiers.
2. The mean IQ obtained by these 100 A W O L's was 92.65. The number of those falling in the "dull normal"

and "borderline defective" categories was about twice the normal expectancy.

3 Comprehension sub-test scores were significantly lower than those of Information, Arithmetic and Similarities. A tentative explanation of this fact was offered: these trainees were deficient in social intelligence and in the ability to properly evaluate past experience.

REFERENCES

1. Cleckley, Hervey. *The Mask of Sanity*. St. Louis: The C. V. Mosby Company, 1941.
2. Wechsler, David. *The Measurement of Adult Intelligence*. Baltimore, Maryland: The Williams & Wilkins Company, 1944.

PREDICTIVE VALUES OF THE STANFORD SCIENTIFIC AND THE ENGINEERING AND PHYSICAL SCIENCE APTITUDE TESTS

H A COOPRIDER

and

H. R. LASLETT

Oregon State College

THE measurement of special aptitudes is a phase of testing that is of unusual importance. Measurements in this field are designed to offer bases for the predictions of probable success in particular fields of human endeavor. Such instruments can have extensive educational value in the control of school admissions and still more in counseling; but evidences of their predictive value for specific purposes must be available if such activities are to be valid.

The problem of this study was the determination of the relationships between scores on two special aptitude tests on the one hand and two scholastic aptitude tests on the other with grade-point averages which had already been earned in several subject-matter fields at the college level. The special aptitude tests were the *Stanford Scientific Aptitude Test* (8) and the *Engineering and Physical Science Aptitude Test* (2). The scholastic aptitude tests were the *American Council on Education Psychological Examination* (5) and the *Ohio State University Psychological Test* (6). The individual scores on these tests were correlated with the individual grade-point averages for courses taken in engineering, mathematics, physical science (excluding chemistry), chemistry, biological sciences, and for all courses taken. The number of academic quarters involved varied between one and six, with an average of slightly more than two.

The number of students involved was 376, all males, who took these various tests during the calendar year of 1946 or the first quarter of 1947. The majority were veterans of the armed services in the recent war. The distribution of the scores of this

group on the A.C.F. Test was parallel to that of the entire student body of 7,500 and the group is held to be a representative sample of that student body.

Major limitations of this study are the low reliability of instructors' marks and the fewness of the quarters available. Rogers (4), however, found an average reliability of college grades for eight semesters of .66, but Traxler (6) found that the reliability of grades varied with the number of terms used. Read (3) found correlations between averages for each semester and the four-year averages of .73 to .83, with a relationship of .76 between the first-semester grades and the four-year averages. Lehmann (1) found that the correlations of first-, third-, and sixth-quarter point-hour ratios, each with final cumulative point-hour ratio, were .66, .80, and .88 respectively. It would appear that grades from two quarters are fairly representative of four-year college scholarship standings.

The *Stanford Scientific Aptitude Test* (8) was copyrighted in 1929 and 1930, and is intended for use with high-school juniors and seniors and with college students. It attempts to measure scientific aptitude by means of eleven sub-tests upon the following topics: experimental bent, clarity of definition, suspended versus snap judgment, reasoning, inconsistencies, caution and thoroughness, discrimination of values in selecting and arranging experimental data, accuracy of interpretation, and accuracy of observation. It is a power examination of the group, pencil-and-paper, and language types. One to two hours is usually required for its completion. It has no set time limit.

The *Engineering and Physical Science Aptitude Test* (2) was copyrighted in 1943, and is intended for use at the college level. It attempts to measure aptitude for engineering and for physical science by means of six individually timed sub-tests on mathematics, formulation, physical science comprehension, arithmetic reasoning, verbal comprehension, and mechanical comprehension. It is a speed test of the group, pencil-and-paper, and language types. It requires a total time of approximately seventy-five minutes for its administration.

The *American Council on Education Psychological Examination* (5) is copyrighted in a new form annually. It attempts to measure general scholastic aptitude at the high-school senior and the college-freshman levels. Its six sub-tests yield two

sub-scores, a Q-score for quantitative comprehension and an L-score for linguistic comprehension, as well as a total or composite score. The six sub-tests are timed separately. It is a group, pencil-and-paper, speed test. It requires approximately sixty-five minutes for its administration.

The *Ohio State University Psychological Test* (6) is copyrighted as it is issued in new form from time to time. It attempts to measure general scholastic aptitude at the high-school senior and college level. It has three sub-tests, the scores for each of which have norms, as well as a total score. It is a group, pencil-and-paper, power test. One to two hours is usually required for its completion. It has no set time limits.

The grades used in this study were all from courses taken at Oregon State College. The grades carry "honor points" in the ratio of four for each hour of "A," three for each hour of "B," two for each hour of "C," one for each hour of "D," and none for each hour of "F." Grades were available for two twelve-week quarters for 62 per cent of the 376 students, for one quarter for 14 per cent, for three quarters for 10 per cent, and for four to six quarters for the remaining 14 per cent.

The various correlations, their probable errors, and the number of cases involved are shown in Table 1. The correlations were found by the product-moment method, and are uncorrected for attenuation.

The results of this study indicate that the best single predictor of total scholastic achievement at the college level among the tests used was either the *American Council on Education Psychological Examination* total score or the *Engineering and Science Aptitude Test* score. Each showed a correlation with total grade-point averages of .51, with a probable error of .03.

Engineering grades could have been predicted equally well from the *American Council on Education Psychological Examination* Q-score or total score, the *Engineering and Physical Science Aptitude Test*, or the *Stanford Scientific Aptitude Test*, with correlations of .40, .39, .39, and .37 respectively. While none of these correlations is high, they are very similar and show that the special aptitude tests were no better tools for prediction than the general aptitude test with this group.

The prediction of mathematics grade-point averages could have been best accomplished by the *American Council on Edu-*

cation *Psychological Examination* Q-score, but it was closely followed by the total score on the same test with a correlation of .41 for the latter in comparison with .43 for the former.

Grades in physical science (excluding chemistry) in this study could have been best predicted by the *Engineering and Physical Science Aptitude Test* with a correlation of .43, but it was followed closely by the *American Council on Education Psychological Examination* (total score) with a correlation of .38.

TABLE 1
Correlations Between Scores on Various Tests and Grade-point Averages in Academic Fields

Grade-point Averages	SSAT	E and PSAT	ACE T	ACE Q	ACE L	Ohio
Total Grade Points.....	.39	.51	.51	.44	.45	.45
P.E.....	.04	.03	.03	.03	.03	.03
N.....	190	308	362	362	362	356
Engineering.....	.37	.39	.39	.40	.29	.23
P.E.....	.06	.04	.04	.04	.04	.04
N.....	82	211	215	215	215	211
Mathematics.....	.33	.35	.41	.43	.31	.26
P.E.....	.05	.03	.03	.03	.03	.04
N.....	160	293	325	325	325	321
Physical Science (excl. chem.)...	.16	.43	.38	.31	.35	.36
P.E.....	.07	.04	.04	.05	.04	.04
N.....	78	173	177	177	177	178
Chemistry.....	.57	.56	.54	.49	.45	.42
P.E.....	.05	.05	.05	.05	.05	.06
N.....	69	71	99	99	99	99
Biology.....	.37	.64	.43	.32	.43	.41
P.E.....	.07	.06	.06	.07	.06	.06
N.....	61	40	79	79	79	76

Success in college chemistry courses could have been predicted quite similarly from the *Stanford Scientific Aptitude Test* scores, the *Engineering and Physical Science Aptitude Test* score, and the *American Council on Education Psychological Examination* (total score) with correlations of .57, .56, and .54.

Biological science grade-point averages could have been predicted most accurately, in this study, from the *Engineering and Physical Science Aptitude Test* scores, with the highest correlation of the study, .64. While the number used in finding this correlation was small the probable error was also small.

Never the less, this particular correlation may well be somewhat fortuitous.

Summary

In this study, the *Stanford Scientific Aptitude Test* scores correlated no better with grades in engineering and science courses than did the *American Council on Education Psychological Examination* partial or total scores, nor did it appear to differentiate between scientific aptitude and general scholastic ability.

The *Engineering and Scientific Aptitude Test*, in this study, did not appear to predict success in engineering and physical science courses appreciably better than did the *American Council on Education Psychological Examination*, but it does appear to correlate slightly better with grades in biological science. In addition, it appears to correlate more closely with science grades than with engineering grades, and with biology grades more closely than other tests designed for that specific purpose.

REFERENCES

- (1) Lehmann, Ruth T. "The Significance of First-Year College Marks." *Educational Research Bulletin*, XXII (1943), 217-218.
- (2) Moore, Bruce V., Lapp, C. J. and Griffin, Charles H. *Manual for the Engineering and Scientific Aptitude Test*. New York: The Psychological Corporation, 1943. Pp. 8.
- (3) Read, Cecil B. "The Relationship of Scholastic Averages in Various Semesters." *School and Society*, LII (1940), 468-469.
- (4) Rogers, Herbert W. "The Reliability of College Grades." *School and Society*, XLV (1937), 758-760.
- (5) Thurstone, L. L. and Thurstone, T. G. *Manual for the American Council on Education Psychological Examination*. Washington: American Council on Education, 1946. Pp. 7.
- (6) Toops, Herbert A. *Manual for the Ohio State University Psychological Test*. Chicago: Science Research Associates, 1941. Pp. 4.
- (7) Traxler, Arthur E. "The Correlation Between Achievement Scores and School Marks in an Independent School for Boys." *Journal of Applied Psychology*, XXIV (1940), 58-63.
- (8) Zyve, David L. *Explanatory Booklet for the Stanford Scientific Aptitude Test*. Stanford University: Stanford University Press, 1930. Pp. 16.

THE CONTRIBUTORS

Sidney Adams—Ph.D., University of California, 1933. Job-description writer, U. S. Employment Service, Occupational Research Program, 1935-1937. Research in trade, clinical and other tests, Tennessee Valley Authority, 1937-1946. World War II service in test research, aviation psychology and clinical psychology, 1941-1945. Test developer, U. S. Civil Service Commission, at present.

Earl F. Carnes—M.A., Ohio State University, 1947. Assistant to the Personnel Director, Antioch College, 1940-1942. Assistant Instructor, Department of Psychology, Ohio State University, 1947-.

Jerry H. Clark—Ph.D., University of Texas, 1948. Assistant in Psychology, University of Texas, 1937-1941. With the U. S. Army and Air Force: Psychological Assistant, Personnel Consultant, Clinical Psychologist, 1942-1946. Instructor in Psychology, University of Texas, Fall 1946. Instructor in Psychology, University of California, Santa Barbara College, Summer 1947. Manager of Guidance Center, University of California Extension, Santa Barbara, California, 1947-.

H. A. Coopridner—Graduate work, Oregon State College, at present. In World War II with a Personnel Replacement depot in the South Pacific, with the Personnel Authorization Division of the Third Service Command in the U. S. Recently associated with the Counseling and Testing Service, Oregon State College. Assistant Registrar in charge of vocational counseling, testing and placement at Aurora College, Aurora, Illinois, at present.

Francis J. DiVesta—M.S., Cornell University, 1946. High-school teacher in Connecticut, 1941-1945, Vice-Principal, 1943-1945. Research Assistant, University Testing Service, Cornell University, 1946-. Author of article in *The Connecticut Teacher*. Member, Phi Delta Kappa.

Frank J. Dudek—Ph.D., University of Southern California, 1947. Aviation Psychologist, AAF Aviation Psychology Program, 1942-1946. National Research Council Predoctoral Fellow, 1946-1947. Assistant Professor of Psychology, Northwestern University, 1947-. Associate Member, American Psychological Association. Member, Midwestern Psychological Association, Phi Beta Kappa, Sigma Xi, Phi Delta Kappa, Phi Kappa Phi, Psi chi.

Henry S. Dyer Ed.D., Harvard University, 1941. Test Technician, National Clerical Ability Tests, 1938-1939. Research Associate, the Graduate Record Examination Project, Carnegie Foundation for the Advancement of Teaching, 1939-1941. Assistant Professor of Psychology, Allegheny College, 1941-1942. Assistant Dean of Harvard College, 1942-1945. Assistant to the Dean of the Faculty of Arts and Sciences, Director of the Office of Tests, Harvard University, 1945-. Member, Psychometric Society. Associate Member, American Association of University Professors.

Daniel D. Feder Ph.D., University of Iowa. Associate, State University of Iowa, 1934-1938. Assistant Director, Personnel Bureau, and Assistant Professor of Psychology, University of Illinois, 1938-1942. Executive Officer and Supervisor, Illinois State Civil Service Commission, 1942-. With the U. S. Navy: Officer in charge of Radio Materiel Unit (formerly Training Activity, later part of Research Activity); Officer in Charge to study German Naval Selection and Training Methods attached to U. S. Naval Technical Mission in Europe. Author of articles on personnel and measurement. Member, American Educational Research Association, American Psychological Association, American College Personnel Association, Civil Service Assembly. President, American College Personnel Association, 1946-1947.

Chester W. Harris Ph.D., University of Chicago, 1946. Teacher, Public Schools, Denver, Colorado, 1935-1939. Research Assistant, Evaluation Staff of the Eight-Year Study, University of Chicago, 1939-1941. Aviation Psychologist, Army Air Forces Aviation Psychology Program, 1942-1946. Assistant Professor of Education and Examiner, University of Chicago, 1946-. Member, Kappa Delta Pi, Phi Delta Kappa. Associate, American Psychological Association.

Herbert R. Laslett-In World War II: Liaison Officer between Fourth Service Command Headquarters, induction stations, reception centers, and State Selective Service Headquarters; Rehabilitation Officer (pardons and paroles) in the European Theater Headquarters.

Milton M. Mandell-B.A., New York University, 1933. Assistant Director of Examinations, Los Angeles City Civil Service Commission, 1939-1940. Classification Consultant, State of Connecticut, 1940-1941. Regional Personnel Officer, O.E.M., 1941-1942. Personnel Officer, Office of Program Vice-Chairman, War Production Board, 1942-1943. Chief Analyst, Committee For Congested Areas, 1943-1944. Chief, Administration and Management Testing, U. S. Civil Service Commission, 1944-. Member, American Society of Public Administration, Civil Service Assembly.

Francis P. Robinson—Ph.D., University of Iowa, 1932. Instructor, University of Iowa, 1932-1933. Chairman, Department of Psychology and Education, Stout Institute, 1932-1937. Assistant Professor of Psychology, 1937-1941, Associate Professor, 1941-1944, Professor, 1944-, Ohio State University. Author of *Diagnostic and Remedial Techniques for Effective Study*. Co-author of *Psychology and the New Education*, and author of articles on student personnel and educational psychology. Fellow, American Psychological Association.

Verner M. Sims—Ph.D., Yale University, 1926. High-school teacher, Kentucky and Connecticut, 1920-1926. Professor of Education and Psychology, Louisiana Polytechnic Institute, 1926-1928. Associate Professor of Psychology, University of Alabama, College of Education, 1928-1939. Professor and Staff member, Bureau of Educational Research, University of Alabama, 1939-. Member, American Psychological Association, American Educational Research Association.

Edmund G. Williamson—Ph.D., University of Minnesota, 1931. Assistant Professor of Psychology, 1930-1938, Associate Professor, 1938-; Director of University Testing Bureau, 1932-1938, Coordinator of Student Personnel Services, 1938-1941, Dean of Students and Professor of Psychology, 1941-, University of Minnesota. Author of books and articles on counseling and personnel work. Member, American Psychological Association, Sigma Xi, Phi Delta Kappa, Psi Chi, Alpha Kappa Lambda, American College Personnel Association, American Association for Applied Psychology, Advisory Committee to the U. S. Armed Forces Institute. On Editorial Board, EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT. Chairman, Committee on Student Personnel Work of American Council on Education. Recently returned from three-month survey of industrial and educational personnel work in Germany.

Asahel D. Woodruff—Ph.D., University of Chicago, 1941. Director, L.D.S. Institute of Religion, Ogden, Utah, 1941-1942. Instructor in Educational Psychology, 1942-1944; Assistant Professor, 1944-1946; Associate Professor, 1946-1947; Director of the University Testing Service and of the Bureau of Educational Research and Service, 1947-, Cornell University. Author of *Psychology of Teaching*, and of articles on human motivation. Member, American Psychological Association, American Association for the Advancement of Science, National Education Association, Phi Delta Kappa.